2020

Interactions Between Humans, Virtual Agent Characters and Virtual Avatars

Tamara Griffith
University of Central Florida

INTERACTIONS BETWEEN HUMANS, VIRTUAL AGENT CHARACTERS AND VIRTUAL AVATARS

by

TAMARA S GRIFFITH
B.S. University of Central Florida, 1992
B.S. University of Central Florida, 2003
M.S. University of Central Florida, 2011

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the School of Modeling, Simulation, and Training
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term
2020

Major Professor: Patricia Bockelman

ABSTRACT

Simulations allow people to experience events as if they were happening in the real world in a way that is safer and less expensive than live training. Despite improvements in realism in simulated environments, one area that still presents a challenge is interpersonal interactions. The subtleties of what makes an interaction rich are difficult to define. We may never fully understand the complexity of human interchanges, however there is value in building on existing research into how individuals react to virtual characters to inform future investments. Virtual characters can either be automated through computational processes, referred to as agents, or controlled by a human, referred to as an avatar. Knowledge of interactions with virtual characters will facilitate the building of simulated characters that support training tasks in a manner that will appropriately engage learners.

Ultimately, the goal is to understand what might cause people to engage or disengage with virtual characters. To answer that question, it is important to establish metrics that would indicate when people believe their interaction partner is real, or has agency. This study makes use of three types of measures: objective, behavioral and self-report. The objective measures were neural, galvanic skin response, and heart rate measures. The behavioral measure was gestures and facial expressions. Surveys provided an opportunity to gain self-report data. The objective of this research study was to determine what metrics could be used during social interactions to achieve the sense of agency in an interactive partner.

The results provide valuable feedback on how users need to see and be seen by their interaction partner to ensure non-verbal cues provide context and additional meaning to the dialog. This study provides insight into areas of future research, offering a foundation of knowledge for further exploration and lessons learned. This can lead to more realistic experiences that open the door to human dimension training.

This is dedicated to my family, who has been patient and supportive with me throughout this process. In addition, I appreciate the never-ending support of my work family, who has also provided support and some of the most amazing opportunities an individual could hope for. Finally, I'd like to express my deepest gratitude to my dissertation committee. Your involvement and support have been priceless.

ACKNOWLEDGMENTS

The completion of this dissertation will be a culmination of many years of work. As with anyone facing this moment, there have been highs and lows. I would never have made it to this place without the unending support from my dissertation committee who invested many hours of their personal time to ensure I was on the right path. Dr. Patricia Bockelman stepped into the role of my committee chair and immediately I felt that I had gained a best friend, albeit one who could tell me honestly when I didn't have my head screwed on right. She was able to ensure research rigor with a firm but kind hand. I have been honored to benefit from her guidance. Dr. Cali Fidopiastis has also been a strong supporter. She offered not just her profound knowledge in neurological measurements, but also her insight into what makes sense in the topics that matter to me. Her strength and confidence come from years of practical experiences in the commercial and research worlds, the value of which cannot be measured. And like icing on the cake, Dr. Joan Johnston, a coworker and friend, has brought her vast knowledge of research design, analysis, and analytical tools. Her guidance cleared my confusion and pulled back the curtains of intimidation. She fostered a passion for examining data that I didn't think lived within me. These women have been so inspiring and so complimentary of one another's styles; I cannot imagine being here without them and I cannot thank them enough for their guidance and friendship.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ABM	Advanced Brain Monitoring
AI	Artificial Intelligence
ANS	Autonomous Nervous System
ASA	Advanced Situational Awareness
ASCOPE	Area, Structures, Capabilities, Organizations, People and Events
ANS	Autonomic Nervous System
AWFC	Army Warfighter Challenges
CAM	Computer-as-Medium
CAS	Computer-as-Source
CREST	Center for Research and Education in Sexual Trauma
CTC	Combat Training Center
DS2A	Digital Survivor of Sexual Assault
EDA	Electrodermal Activity
EEG	Electroencephalography
ERP	Event Related Potentials
fMRI	Functional Magnetic Resonance Imaging
GSR	Galvanic Skin Response
HUMINT	Human Intelligence
HQ	Headquarters
IBI	Inter (heart) Beat Interval
ICT	Institute for Creative Technologies
IEWTPT	Intelligence Electronic Warfare Tactical Proficiency Trainer
M&S	Modeling and Simulation
METT-TC	Mission, Enemy, Terrain and weather, Troops and Support Available, Time Available, and Civil Considerations
OODA	Observe, Orient, Decide, Act loop
PET	Positron Emission Tomography
PD	Prisoner's Dilemma
PMESII-PT	Political, Military, Economic, Social, Information, Infrastructure, Physical Environment and Time
PTSD	Post-Traumatic Stress Disorder
SCR	Skin Conductance Response
SITREP	Situation Report
SPIN	Social Phobia Inventory
SRCT	Social Response to Communication Technology
STTC	Simulation Training & Technology Center, U.S Army Modeling and Simulation Research
STC	Superior Temporal Sulcus
ToM	Theory of Mind
U.S.	United States

USC	University of Southern California
VA	Virtual Agent
VMPFC	Ventromedial prefrontal cortex

CHAPTER ONE: BACKGROUND

Since before written record, humans have simulated various events and used models to represent the players (U.S. Army Modeling and Simulation Office, 2011). Improvements in computation and network performance has led to improvements in the realism of Modeling and Simulation (M&S) applications.   M&S tends to cost less and be safer than live training (Smith, 2010).   Now it is possible to place a user in the action in a way that mimics the real world (Dukstein, Watkins, & Deakins, 2007), and the range of applications continues to expand.

With the wide range of applications that make use of M&S tools, the entertainment industry provides a pervasive example that blurs the lines between live action and special effects. The commercial world has made significant advances in voice recognition and Artificial Intelligence (AI) to support marketing and automated service calls, sometimes causing confusion about whether we are speaking to a human or AI. While the topic is relevant to the general population, the focus of this research is how it might ultimately support training.

M&S is used throughout virtual simulations to support various tasks in fields such as: cyber, medical, evaluation, training and education (U.S. Army Modeling and Simulation Office, 2016). An example of a training application is the Intelligence Electronic Warfare Tactical Proficiency Trainer (IEWTPT). This tool includes a Human Intelligence (HUMINT) component that uses life-like avatars and speech recognition to support tactical questioning, interrogation, screening, and the use of an interpreter through free-flowing conversations with virtual humans (Blinde, 2016).

M&S is used to practice dangerous tasks and conduct analysis, as well as being a cost-avoidance strategy in the place of live training (Bukhari, Andreatta, Goldiez, & Rabelo, 2017).   Every deploying soldier has experienced some type of M&S training to develop critical warfighting skills (U.S. Army Modeling and Simulation Office, 2016). M&S holds a critical role in ensuring that the U.S. Army remains prepared as a premier combat force (U.S. Army Modeling and Simulation Office, 2016). This means the U.S. Army must continue to invest in, and improve, its M&S capabilities.

To keep current with evolving technology, the U.S. Army has exploited advances in the commercial game and virtual world communities. While traditional Army training systems can be very costly, commercial game developers spread development costs across millions of users, keeping end-product costs low. Despite that, some commercial games are capable of simulating synthetic environments at a higher fidelity than some military training simulations (Alexander, Brunye, Sidman, & Weil, 2005). While the military and commercial worlds are driven by different motivating factors, they can be complementary. For example, commercial game developers want to completely immerse their users in the game to keep them coming back for more. Likewise, military training that is engaging may bring about better training outcomes (Schobel, Janson, Hopp, & Leimeister, 2019); therefore, engagement is another area where virtual simulations can benefit from commercial games.

Simulated experiences that are effectively engaging and immersive may involve the user so completely that (s)he will experience stress, fear, excitement or anger as a result of the unfolding events (Maxwell, Griffith, & Finkelstein, 2014). There is minimal physiological, emotional and physical difference between the immersive environment and the real world (Blascovich, et al., 2002). Keeping the user engaged in the environment is important to reach a state of "flow". Flow is a mental state where a person feels fully immersed in what (s)he is doing to the point of actually gaining energy and losing track of time (Csikszentmihalyi, 1998). This theory is a cornerstone to current-day game design and is the focus of a great deal of learning research (Chang, Liang, Chou, & Lin , 2017) (Pearce, 2005) (Shernoff, Csikszentmihalyi, Schneider, & Shernoff, 2014).

While engagement and flow are important to retain the attention of learners the focus must always be on the learning objectives. This is an important difference between commercial games and applications used for training and learning. The goal is to ensure that investments in realism directly support training tasks as sometimes awe-inspiring realism can actually distract the learner from the learning material (Aitamurto, Boin, Chen, Cherif, & Shridhar, 2018).

Interpersonal Interactions

Despite the vast improvements in realism in simulated and game environments one area that still presents a challenge is simulated interpersonal interactions. Many games use multiple choice options and cinematics to give the sense of realism to interactions, but it can minimize the sense of control the user experiences. This is a cost-effective strategy to push the story line forward in the commercial game world, but it can change the player to a passive watcher as the interchange unfolds. The challenge is in maintaining user engagement involving the human dimension of M&S experiences.

Human interactions are complex and varied. One can communicate volumes through a gesture or very little through dialog. The subtleties of what makes an interaction rich are difficult to define. Events in an interaction can cause people to shut down and refuse further engagement. While we may never fully understand the complexity of human interchanges, there is value in building on existing research into how we react to different communication modalities and virtual characters that are either fully automated (which is considered a computational agent) or controlled by a human (referred to as an avatar). This understanding will inform future investments in virtual characters by helping to understand what drives people to engage or disengage. This knowledge will facilitate the building of simulated characters and avatars that will support training tasks in a manner that will appropriately engage learners.

To better understand interpersonal interaction in M&S, it is useful to explore a similar industry. Online marketing and service calls often use an automated agent to sort calls then transition from the agent to the appropriate human. A good example of this is when you have an issue with a bill. You contact the company and are connected to an automated agent who is able to interpret basic verbal dialog. The agent might collect your name, phone number and account number along with some rudimentary information about the issue. If the agent is unable to satisfy your request, the call is transferred to a live operator who benefits from the information collected by the agent (Pradana, Goh, & Kumar, 2018). This process optimizes the live operator's time, allowing them to focus on tasks that require human intelligence. Leveraging this commercial work, the U.S. Army can use M&S to render realistic and cost-effective training.

For example, consider a course on Advanced Situational Awareness (ASA). During this course, soldiers perform surveillance on a village from a distance to look for patterns and changes over time. They may also enter the village and talk with key leaders. During face-to-face interactions soldiers need to be aware of subtle physiological changes that may indicate stress, fear or dishonesty.   The U.S. Army currently hires live actors to play the role of villagers within a mockup of a small village. They demonstrate normal patterns of life, react to soldier's presence and effect complex social interactions in face-to-face interchanges. This training is critical but carries a high manpower cost. One way to drive down cost, while maintaining the training value, is to conduct the training in an M&S environment. Then, in much the way call centers have automated mundane actions, artificial intelligence agents can demonstrate patterns of life and even disruptions in those patterns. Detailed interpersonal tasks can be performed by human-controlled characters, or avatars. Actors control the avatar's facial expressions, eye movement and gestures as easily as they manage their own natural body movement. A small number of actors control multiple avatars by jumping from one character to the next as the scenario requires. This functionality provides an equivalent level of training at a fraction of the price. Live training may still be desired after the virtual training, but the duration and cost could be significantly reduced. The following sections describe the technology needed to pursue training of this type.

Virtual Characters

While the capabilities of agents (driven by artificial intelligence) and avatars (driven by humans) have improved significantly in recent years, there is still a significant gap between the realism of interactions with humans and interactions with an agent. There has been a great deal of research into this phenomenon (Blascovich, et al., 2002), (de Melo & Gratch, 2015) , (Pena, Khan, & Alexopoulos, 2016), (Heyselaar, Hagoort, & Segaert, 2017), but it is still unclear what it is about computer agents that might cause disengagement during the interaction. This research will push toward a deeper understanding of the differences between various modalities of interaction to inform both Army training and the existing body of literature on the subject, but first it is important to explore the difference between an agent and avatar along with existing research into those differences.

Theories on Virtual Character Interactions

The Social Response to Communication Technology

The Social Response to Communication Technology (SRCT) was a popular theory in the 1990s that posited humans evolved assuming that every person was a real person. As such, automatic responses were developed that simplified cognitive parsing (Morkes, Kernal, & Nass, 1999) (Reeves & Nass, 1996). SRCT is built on Spinoza's (Deleuze, 1988) theological-political treatise that acceptance of an idea is (1) part of the automatic comprehension of that idea and (2) the rejection that an idea occurs subsequent to, and uses more effort than, its acceptance (Gilbert, 1991). An example of this concept is when people automatically thank the technology for assistance. Verbal interactions with technology still lack the natural fluidity of human-to-human interaction, so the sense of realism often falls short (Branigan, Pickering, Pearson, & McLean, 2010). As research was conducted in this area, focusing on verbal realism (Nass & Lee, 2001), social behavior (Nass & Moon, 2000) and behavioral realism (Blascovich J. , 2002), the theoretical constructs evolved. This evolution is described below.

The Ethopoeia Concept

Based on the importance of social cues, Nass and Moon (2000) developed the Ethopoeia Concept. This suggests that as long as the situation includes social cues, such as interactivity, natural speech, or the filling of social roles, social scripts trigger automatic social behavior. This is exemplified in a study (Appel, von der Putten, Kramer, & Gratch, 2012) that found participants who stated explicitly that treating a computer like a human is not appropriate still interacted in a social way with agents as long as they believed they were interacting with a human. This concept did not identify the characteristics of the interpersonal interactions in an M&S environment that drove people to react socially to the agents, so further exploration followed (Nass & Moon, 2000) and led to the Revised Ethopoeia Concept which states that more human-like characteristics or more behavioral realism, will lead to more social reactions by the user. Some researchers also wondered about the importance of physical realism and agency. Agency refers to the belief that the interaction partner is a real human (Appel, von der Putten, Kramer, & Gratch, 2012). As research continued on this topic, the question became: how much of each of these factors determine if an individual will engage with a virtual character as if it had agency. This question led Appel et al. (2012), to propose the Threshold Model of Social Influence.

The Threshold Model of Social Influence

The Threshold Model of Social Influence states that a human being initially only responds socially to another human being (Appel, von der Putten, Kramer, & Gratch, 2012). In a virtual environment a human will respond socially to a virtual character only if the behavior is so realistic that the user cannot distinguish a computer-controlled agent from a human-controlled avatar (Blascovich J. , 2002). Within this concept social verification is used to assess if engagements are semantically meaningful during communication with virtual others. Behavioral realism and agency are two factors humans apply for social verification which are perceived on a low to high continuum. The participant responds to the character socially once a threshold of social influence has been crossed. Behavioral realism is the level at which the character acts and looks real (Blascovich J. , 2002). Character agency is the sense that the character can think, has a history, can plan and act, and has opinions (Blascovich J. , 2002). An additional factor, experience, has also been explored, which is the character's ability to sense and feel emotion (de Melo & Gratch, 2015).

The evolution of these concepts has helped us understand interactions better, but determining where the threshold is for each factor and what factors are more important than others still needs further research. The study described in this dissertation compares cognitive markers in interactions with various modes of communication (face-to-face, text, and video) and with different interactive partners (person, human-controlled virtual character, and algorithm-controlled character). Ultimately, significant variations in cognitive markers could help establish an actual threshold. This threshold might be based on the communication mode, or the interactive partner's ability to portray character agency. It is possible that the threshold will be crossed when the interactor demonstrates opinions or the ability to sense emotion by the participant. The next section presents an application where virtual character realism might be important.

Army Training Applications

The work in this dissertation focuses on M&S learning tools and the factors that set human, avatar and agent interactions apart in the areas of engaging participants to supporting training. Research conducted by Brimstin, Higgs and Wolf (2015) and Milham, et al., (2017) on stress inoculation during Army training, has shown that a high resolution game environment with human-controlled-characters that depict facial expressions, enabled soldiers to observe Advanced Situational Awareness (ASA) cues. These cues were used to identify when a person was being truthful, evasive, or lying (Brimstin, et al., 2015). The research indicated that human-controlled-characters had crossed the threshold into believability and supported the training task of observing and responding to ASA cues. This research used surveys from subject matter experts, instructors and trainees (Milham, et al., 2017). Further research into this learning task could make use of quantitative and physiological measures to help establish what factors are most important to bring about stress inoculation and ASA. The research described below is intended as a next step to reach this ultimate goal.

Research Problem

Ultimately, the goal is to understand what makes virtual characters appear to have agency. To answer that question, it was important to establish metrics that would indicate when people believe their interaction partner had agency. This study made use of three types of measures: objective, behavioral and self-report. The objective measures were neural measures, Galvanic Skin Response (GSR) and heart rate measures. The behavioral measure was gestures and facial expressions. Surveys provided an opportunity to gain self-report data. The objective of this research study was to determine what metrics could be used during social interactions to understand the sense of agency in an interactive partner.

Biometric Measures of Social Interactions

This research made use of three types of biometric measures. Electroencephalography (EEG) provided the neural measures. Electrodermal Activity (EDA) used a GSR sensor and a heart rate sensor collected heart rate variability. Each of these measures and related research is described below.

Neural Measures

Neural measures have a history of being applied to understand social interactions. For example, neural measures captured while playing social games provide insight into the bio-mechanisms of interaction. Sanfey (2007) explored games as a measure of social decision making as it related to social reward, competition, cooperation, and coordination as well as strategic reasoning. Certain portions of the brain activate as a result of social rewards (Cromwell, Tremblay, & Schultz, 2013) (Ascoli, 2014). Emotions are often triggered in social interactions and as such, can be an indicator of character realism in a communication partner. Neural measures of brain activation in areas representing emotion may indicate that the participant views the social partner as having passed the threshold of realism. Further information on the use of EEG in the area of social games can be found in CHAPTER TWO.

Researchers (Ismail, Hanif, Mohamed, Hamzah, & Rizman, 2016) have mapped different brainwaves to emotions. For example, delta waves are associated with sleep and fatigue. Theta waves are associated with stress relief and memory recollection. There may also be a relationship between theta and emotions. Alpha waves are related to relaxation, creativity, and visualization. There is a relationship between alpha waves, reflection, and problem-solving. Beta waves are linked to alertness, logic, and concentration. There is a relationship between beta and productivity. Gamma waves can be linked to complex tasks, learning, information processing, and ideation. The roles these brainwaves play in this research is further explored in the Discussion section.

Electrodermal Activity

When an individual's sympathetic nervous system activates, it causes the release of small amounts of sweat, which cause changes in skin conductance (Neulog , 2018). EDA has been associated with measures of arousal (Egan, et al., 2016) (Drachen, Nacke, Yannakakis, & Pedersen, 2010). Prior research (Paiva, 2000) (Wilson & Sasse, 2000) shows that EDA is a reliable measure of stress as well. Arousal/Stress during a social interaction may indicate that the interaction partner is perceived to be another human. The intent of EDA in this study is to assess whether variability in EDA supports EEG data. If there is a strong correlation, it might be possible to use EDA in the place of EEG data in future research.

Heart Rate Data

Heart rate variability measures Autonomic Nervous System (ANS) activity, which is associated with stress (HRV Course, 2016). A study by Hartanto et al. (2014), used heart rate as one indicator of stress when exposing individuals with social anxiety to virtual characters in stress-inducing situations. Researchers were able to manipulate stress in participants based on dialog choices and apparent emotion in the character they were interacting with. Chalmers et al. (2014), found that reduced heart rate variability was linked with social anxiety through their meta-analysis. Similar to EDA, the goal of using heart rate data in this research is to explore whether stress, as indicated by heart rate variability, might be an indicator of whether the virtual character has exceeded the threshold of realism required for the participant to feel they are engaging with an actual person.

Behavioral Measures

Behavioral measures were also considered, specifically the number of gestures and distinct facial expressions each dialog partner used during their interchange. These gestures might be an indicator of having reached the threshold of conversational realism. Multiple researchers have explored how the number of gestures varies based on the visibility of the listener (Alibali, Heath, & Myers, 2001) (Ozyurek, 2002) (Krauss, Dushay, Chen, & Rauscher, 1995) (Rime, 1982). Jacobs and Garnham (2006) designed research that strongly supports the primarily communicative function of gestures. Their research indicated that "more gestures were produced when the listener appeared attentive than when the listener appeared inattentive," and that "speakers adapt their gesture usage to the perceived requirements of the listener" (Jacobs & Garnham, 2006). Their research demonstrated that gestures occur to benefit the listener. If gesturing is a tool used to bolster communication, it can provide insight into variations in communication strategies and communication partners.

Objectives

This study sought to examine interpersonal interactions in an M&S environment through the examination of cognitive markers during verbal and nonverbal interactions between two people communicating: face-to-face; via video teleconference; via a human-controlled avatar; via a computer-controlled character; and via text messaging. The outcome can inform future investments in virtual characters to support training tasks.

The scenario was intended to evoke an emotional response and empathy/sympathy with the character during the interaction. A virtual character that exceeded the threshold necessary to convince an interaction partner that they were real could support various types of training within the Army and across other services. For example, the Defense Equal Opportunity Management Institute (DEOMI) is exploring this technology to train human resource personnel to recognize indications of sexual harassment, Post-Traumatic Stress Disorder (PTSD) and depression. In addition, the U.S. Army models realistic, sympathetic characters in military simulations for stress inoculation, to improve rapport-building skills and to depict body posture and movements with the long-term goal of portraying characteristics such as dishonesty, fear, friendliness and distrust.

This is a mixed-methods exploratory design. Research Questions follow:

Research Questions

What metrics can be applied as an indicator of perceived agency?

1. *Biometric Correlates of Social Interactions*

Research Question 1: Are Biometric Correlates of Social Interactions appropriate to measure the level of perceived agency based on study condition?

Research Question #1a – Variations in EEG Data

Do variations of brain activity (engagement, wavelength, and workload) in the different study conditions indicate that participants respond to communication mode and/or the interaction partner in differing ways?

Research Question #1b – Variations in Electrodermal Activity

Does arousal, as measured by EDA, indicate that an interaction partner has perceived agency?

Research Question #1c – Variations in Heart Rate Data

Does the heart's Inter -Beat Interval provide a measure that indicates that an interaction partner has perceived agency?

2. *Behavioral Measures*

Research Question 2: Are behavioral measures appropriate to assess perceived agency based on study condition?

3. *Survey Measures:*

<u>Research Question 3:</u> Does survey data indicate variations in perceived agency based on condition?

Research Question #3a – Presence Questionnaire

Will the Presence Questionnaire (APPENDIX C) provide a measure of perceived agency based on condition?

Research Question #3b – Rapport Questionnaire

Will the Rapport Questionnaire (APPENDIX A) provide a measure of perceived agency based on condition?

Research Question #3c – Interaction Questionnaire

Will the Interaction Questionnaire (APPENDIX B) provide a measure of perceived agency based on condition?

Research Question #3d – Social Phobia Inventory (SPIN)

The Social Phobia Inventory (APPENDIX D) measures the severity of social anxiety disorder. Will individuals with Social Phobia, as indicated by the results of the SPIN show different results in the above measures as compared to individuals who don't have indications of Social Phobia?

Design and Analyses. The study has five conditions.

The face-to-face encounter demonstrated the gold standard of the conditions. In each category the face-to-face measurement was expected to provide an example indicating that the interaction partner had exceeded the threshold of realism and a sense of agency to the participant.

The video interaction was expected to be similar to the face-to-face interaction since both the actor and the participant were able to perceive immediate response, feedback, and synchronized gestures in this condition.

For each of the rest of the conditions, the measures were expected to degrade in different ways. Those degradations were expected to provide insight into degradation in other simulated or virtual experiences either with agents or with avatar characters.

Text interaction is common in our current world. It allows communication with one or more people but does not provide visibility into their outward expressions of mood. It is the inability to express emotion during a text interchange that has likely supported the growth in emojis that attempt to provide that context. Workload was expected to change due to the act of typing, depending on the proficiency of both interaction partners.

Parameters for each of the conditions were compared to the face-to-face data to determine what differences appear in each condition. The results were recorded in a chart.

The resulting comparison between the modes of communication was expected to provide valuable information into how humans engage cognitively with other individuals, and non-human interlocutors to inform investments in modeling and simulation.

Potential Ethical Issues

The dialog topic selected for this research was driven by the artificial intelligence functionality available at the time this research was proposed. Specifically, the realistic interactive dialog of the Digital

Survivor of Sexual Assault (DS2A) application developed by the University of Southern California (USC) Institute for Creative Technology (ICT). It used the same technology as the Virtual Survivor Visualization, New Dimensions in Testimony which shows a hologram of a holocaust survivor responding to questions about their personal experiences. DS2A is not a hologram but is rather shown on a computer monitor. It is an application that introduces participants to Specialist Jarett Wright. He is an actual survivor of sexual assault that occurred while he was in the Army. Jarett responds to various questions from an individual via voice recognition. He is a photorealistic video representation controlled through voice parsing and complex branching. The actor representing Specialist Wright in the other conditions used information from Specialist Jarett Wright's depiction in the D2SA application to shape the story he shared. Each participant engaged in 8-10 minutes of dialog about the sexual assault incident driven by questions chosen and asked by the participant through natural dialog. Since there was no way to know how sensitive the topic was to the participant, the proctor received rudimentary training to help the participant unpack the experience and discuss it afterwards. This was done through a task where the participant was asked about the interview. If there were indications that the participant was suffering from a reaction to the content, the proctor provided contact information to the University of Central Florida (UCF) Counseling and Psychological Services (CAPS) and/or the Center for Research and Education in Sexual Trauma (CREST) in the UCF RESTORES lab. The RESTORES lab agreed to provide counseling support to any participants in need. Additionally, resources such as the sexual assault hotline were available to participants who found that they were negatively affected by the topic.

Mixed Methods Study

This mixed methods study addressed the cognitive cues of social interactions to better understand how an individual perceives social information. This was a concurrent, convergent, mixed-methods design. In this design, qualitative and quantitative data were gathered at the same time, but separately from one another, analyzed separately, then the findings were compared. This served to validate both the qualitative and quantitative findings (if the results from the two methods agreed), or to generate insights into the need of further research (if the results from the two methods disagreed). In this study, biometric correlates of social interaction were used to assess the research question involving differences in biometric measures across conditions. Survey data explored user's perception of the social interaction and what factors, if any, drove them to disengage from the interaction.

This research made use of the B-Alert X10 and B-Alert X24 (EEG) Headset System by Advanced Brain Monitoring. The system functions wirelessly through Bluetooth signal transmission. It detects delta, theta, alpha, beta, gamma and high gamma frequencies (Advanced Brain Monitoring, Inc, 2015). The system measures activations of various areas of the brain at different ranges of frequencies. These activations were synchronized with the activities of the participant. These events were then compared across each condition to explore difference. These data helped to determine what brain events were activated in the various social interaction conditions of the study.

CHAPTER TWO: LITERATURE REVIEW

Modeling and Simulation

This study sought to explore interpersonal interactions in an M&S environment through the examination of various measures. The measures help us understand participant's perceptions about the agency of their dialog partner.

To better understand agency and its role in M&S, it is useful to explore how M&S tools are used. M&S tools support a wide range of tasks (U.S. Army Modeling and Simulation Office, 2011). For example, M&S is used to practice dangerous tasks, conduct analysis, and as cost-avoidance for live training (Bukhari, Andreatta, Goldiez, & Rabelo, 2017). One strategy to reduce the cost and development timelines for M&S tools is to capitalize on advances in the current commercial game and virtual world communities (Smith, 2010).

The development of engaging games is a multi-billion-dollar industry (Ricard & Warzynski, 2014). In this industry, cognitive engagement can roughly be tied to revenue. There are sub-factors that also influence success, such as the sense of presence, absorption, flow, immersion and involvement (Li, Jiang, Tan, & Wei, 2014). Factors that can make commercial games successful, can also be applied to learning tasks as long as the design directly supports the learning tasks (Ritterfeld, Cody, & Vorderer, 2009).

Simulated environments are often populated with virtual characters. These characters might be employed in an environment to give the impression of realism with people moving about or engaging in various activities. The characters might provide information on the user's task in the simulated space or could give background to set the stage for events that will follow. Whatever their role, they are an important component of most simulated spaces and their representation can come in the form of avatars and agents.

Avatars and Agents

A human representation within a virtual environment can be an agent or an avatar (Fox, et al., 2015). The difference lies in the locus of control; agent's actions are controlled by algorithms, while avatars are controlled by a human (de Melo & Gratch, 2015).

Artificial Intelligence/Agent

Agents are virtual characters whose actions are controlled by computer algorithms that give the sense that the agent is behaving in an intelligent way. Artificial intelligence (AI) is described by Garnham (2017) as the science of thinking machines, although the actual mechanisms of thought are not necessary to reproduce human-like behavior. In fact, AI is often programmed as a set of decision trees with various levels of branching complexity (Russell & Norvig, 2016). Advances in deep learning, through neural network-style models have improved a wide range of domains associated with AI (Lake, Ullman, Tenenbaum, & Gershman, 2017). Traditional computational paradigms were based on a clearly defined set of instructions (logic), however, neural networks or artificial neural computation draws its methods from statistical physics, with the system "learning", or improving performance by considering the probability of choices based on examples (Hertz, Krogh, & Palmer, 2018).

AI, and associated neural networks, is becoming more prevalent in today's society. In fact, both the terms Artificial Intelligence and neural networks have grown and expanded to touch a wide range of topical areas, such as; pattern recognition, medical diagnosis, sonar signal processing, fault diagnosis, robotics, marketing, financial analysis, data collection, and data fusion (Maren, Harston, & Pap, 1990). While the focus of this paper is on conversational agents, neural networks have also significantly improved speech recognition which makes conversational agents more plausible. The use of voice to interact with an agent is best described in the following scenarios.

Consider the scenario of a call to the phone company and the ensuing interactions with the conversational agent or artificial intelligence character. The agent answers the call, asks for information and attempts to either resolve the issue or sort the call to the appropriate live agent (Pradana, Goh, & Kumar, 2018). This is occurring with increased frequency on websites as well. If one were to linger on a site, a chat box opens asking if help is needed. Rudimentary information is collected, via a text dialog, until a live agent becomes available (Pradana, Goh, & Kumar, 2018). Examples of the use of intelligent agents include; telemarketing (Mor, Cortez, & Rita, 2018), banking (Lacity, Willcocks, & Craig, 2017), interactive marketing (Pradana, Goh, & Kumar, 2018), and personal assistants (Berry, et al., 2017). Sometimes these transitions from AI to live agents are seamless. Sometimes the user gets frustrated and disconnects. This may lead to poor reviews or loss of sales. As such, commercial industry has made significant investments into the technology to increase realism and maintain the connection (Pradana, Goh, & Kumar, 2018). These investments can also benefit the development of artificial intelligence characters, or agents, within games and M&S environments.

These agents have a variety of roles and tasks. For example, they may provide a quest, impart information, act as an opposing force, build a story line or provide comic relief (Millington & Funge, 2009). While some commercial games present objectives to the player via a pop-up display or radio communication, quests are often provided by a Quest Giver (Grey & Bryson, 2011). Some games and virtual experiences make use of a conversational bot. A conversational bot or chatbot is an implementation of Artificial Intelligence which users can interact with by text or voice conversations through software or an application (Russell & Norvig, 2016). Quite often game developers provide a limited number of dialog options or branches rather than allowing free dialog as a strategy to focus the conversation and avoid player frustration. This is because, even with current voice recognition technology, players can get frustrated and quit the game if they cannot immediately determine the right phrasing to elicit the desired AI response (Webster, 2016). This is very different than interactions with other humans who consider context and social norms when responding to queries. This contextual dialog can only currently be provided through a human-controlled avatar.

Avatar

An avatar is a virtual representation of a human that is controlled by a human (Lim & Reeves, 2010). Human control can be as simple as navigation using a keyboard and mouse or as complex as mapping facial and body gestures onto the character. The latter example is similar to controlling a puppet and can best be described through Neal Stephenson's (1995) book, The Diamond Age: or, a Young Lady's Illustrated Primer. The story describes an interactive book that provides an education to a young girl. The book automatically provides educational material to the girl based on her unique needs, but there are some needs automation cannot meet. For this, Stephenson created "ractors" these are actors in interactive movies (Stephenson, 1995). The ractors receive guidance on their role similar to character notes in a script, which they use to "become" the character. This takes place in a booth where the ractor's facial expressions and gestures are mapped onto the character. One ractor can interact with multiple individuals all over the globe. In this capacity, Stephenson's ractor is able to complete the young girl's education where automation falls short. Even science fiction describing a distant future considers the limitations of automation and the need for humans to play the role of humans. Yet, it is not clear exactly what it is about human interactions that automation cannot achieve. This question is fundamental to the research described in this dissertation and is important in determining just how virtual characters can support training tasks.

Virtual Characters in Training

The research described in this dissertation explores the role virtual characters hold for learning tasks in support of Army training. Previous research by Milham, et al., (2017) on the topic, indicates that improved training methodologies and technologies can develop cognitive skills and mental resilience at the squad level by establishing more realistic combat exercises. This can be accomplished through early and continuous Stress Exposure Training (SET) to inoculate against Post-Traumatic Stress Disorder (PTSD) and better prepare soldiers for the stressful situations that are a natural part of combat operations. The study showed that "in a high resolution game environment, with detailed character facial expressions, soldiers were able to observe Advanced Situational Awareness (ASA) cues that identify when a person was being truthful, evasive or lying" (Milham, et al., 2017). The subject squad stated that the training was "awesome, because this was serious dialog…everyone was taking this very seriously…at no point did I feel like this was a check-a-box type training…I was 100% immersed in what was going on." (Milham, et al., 2017). Trainees stated that at one point a virtual character became very agitated while being questioned, providing valuable biometric and kinesics ASA feedback and providing an excellent instructional moment for the subject matter experts during the after action review (Brimstin, et al., 2015). Kinesics is defined as "a systematic study of the relationship between nonlinguistic body motions (such as blushes, shrugs, or eye movement) and communication" (Merriam-Webster, 2019). The research relied on trainee, trainer and subject matter expert surveys to establish the training value of the avatars.

The U.S. Army Combined Arms Center established the Human Dimension White Paper: A Framework for Optimizing Human Performance (2014) which defined the strategy to respond to "a competitive environment that challenges US interests." The report cites a "complex and dynamic mix of cultures and a broad range of actors" as factors that lead to regional instability and conflict. The document stresses the importance of cultural understanding with a nuanced appreciation of social context balanced with ethical and strategically appropriate responses that emphasize cultural empathy and social intuition for the operational environment (Page 7). The Army Warfighting Challenge (AWFC) Framework, as defined in the Training and Doctrine Command Pamphlet 525-3-1, The Army Operating Concept (2014) states that its primary AWFC is to "Develop Situational Understanding." U.S. Army Pamphlet TC 7-102, The Operational Environment and Army Learning (HQ Department of the Army,

2014) uses the operational variable of Political, Military, Economic, Social, Information, Infrastructure, Physical Environment and Time (PMESII-PT) as one strategy to build situational understanding of the operational environment. Another useful paradigm to understand the operational environment is through the civil considerations of: Area, Structures, Capabilities, Organizations, People and Events (ASCOPE) (Headquarters, Department of the Army, 2015) and mission variables: Mission, Enemy, Terrain and weather, Troops and Support Available, Time Available, and Civil Considerations (METT-TC). These tools are used to develop the situation understanding that is the focus of the AWFC.

The most realistic training in the Army is capstone training that takes place at the Combat Training Centers (CTC). This follows about a year of preparation for two-weeks of boots-on-the-ground live training (Headquarters, Department of the Army, 2018). One application of the technology described in this paper is the use of both AI and human-controlled avatars that can prepare trainees for a CTC rotation in building their intelligence portfolio. Soldiers can begin getting Situational Reports (SITREPS) that lead to intelligence collection and video surveillance within a virtual representation of the CTC. They can also conduct interviews with real-time human-controlled avatars to collect HUMINT. These experiences can help the soldiers to build their PMESII-PT, ASCOPE and METT-TC portfolios well before they ever set foot at the CTC. This is effectively home station training that sets the stage for their live training event. The soldiers can conduct key leader engagements, interact with the local population through interpreters and build advanced situational awareness skills all within a virtual mockup of the CTC well before their live training event occurs. This expands the training window and optimizes the time at the live range. This is an example of a training scenario that can directly benefit from the research described in this study.

The intent of the research described in this paper is to apply physiological/biometric and behavioral measures along with self-report data to explore participants' perceived sense of agency of virtual characters. This research is expected to add to the existing literature by exploring what factors of realism are important in engaging participants. A description of existing research on the topic of character realism follows.

Character Realism Theories

Theories related to what factors encourage people to interact socially with computers or agents have evolved over the years as technology has improved and will likely continue to evolve over time. The theories below are presented in the chronological order in which they were developed.

Social Response to Communication Technology (SRCT)

The Social Response to Communication Technology (SRCT) is a paradigm describing an individual's social responses toward computers (Nass & Moon, 2000). Study participants in research on SRCT demonstrated behaviors such as: providing better assessments to a computer asking for an appraisal on itself than when the request comes from another computer (Nass, Moon, & Carney, 1999), gender stereotypes based on the gender of the computer voice (Lee, Nass, & Brave, 2000), social rules (Moon & Nass, 1996) and self-disclosure (Moon, 1998). There are multiple explanations as to why individuals apply social rules to computers. Winograd & Flores (1987) suggest that individuals are unaware that it is inappropriate to apply social rules to computers, while Turkle (1984) suggests it may have more to do with youth or socioemotional limitations. Individuals relate to computers as humans even when they have extensive experience with computers and when explicitly aware that treating computers socially is inappropriate (Nass & Moon, 2000), suggesting that ignorance and youth or socioemotional limitations are not valid explanations of the occurrence. As such, two contradictory models; Computer-as-Medium (CAM) and Computer-as-Source (CAS) emerged.

Computer-as-Medium (CAM)

In this model of SRCT, computers function as media through which the user communicates with another individual. This is similar to conversing with the news anchor while watching the news on television (Heider & Simmel, 1944). Traditionally, this model describes the users' psychosocial attributions as being focused on the computer programmer rather than the computer itself (Sundar S. S., 1993). As such, individuals apply social rules and attributes to computers while thinking that the computer functions as a mediator between the programmer (source) and the human. This is exemplified in the experiment by Heider and Simmel (1944) where participants describe the movement of geometric objects on the screen as if the objects themselves had their own motivations and intentions. Since this is what the creator of the image wanted the participant to experience, the participants were orienting not only to the screen and objects, but also to the human creator, despite the fact that the participants denied thinking about the programmer at all (Heider & Simmel, 1944). However, Sundar & Nass (2000) conducted research that provided strong evidence against the CAM model, suggesting instead that participants respond directly to the Computer-as-Source as described below.

Computer-as-Source (CAS)

The Computer-as-Source (CAS) model of the SRCT states that individuals are responding directly to the computer rather than through the computer to another individual. The important element of this model is that individuals respond to characteristics that are fundamentally human (Sundar & Nass, 2000). Studies have demonstrated certain cues a computer relays that evoke social responses. For example, Turkle (1984) demonstrated the importance of language. Meanwhile, Rafaeli, (1985) found that interactivity was a critical factor. Reeves & Nass (1996) suggested that when faced with human-like cues, individuals "automatically and mindlessly" apply a schema associated with human-to-human interactions. They argued that humans have not evolved to differentiate between non-human actors. Individuals prefer to use cognitive shortcuts for social rules rather than apply effortful systematic processing (Sundar & Nass, 2000).

Ethopoeia

Building on the CAM model, Ethopoeia refers to the “immediate automatic and unconscious reaction” to seemingly social characteristics of computers (Reeves & Nass, 1996) (Nass, Tauber, & Reeder, 1993). Some researchers explored mindlessness, which is the idea that humans sometimes function like automatons and treat information in a very fixed way, as if it were true no matter the circumstances (Langer, 1992) (Langer & Moldoveanu, 2000). Others similarly considered that individuals fail to draw novel distinctions between human and computer interactions (Appel, von der Putten, Kramer, & Gratch, 2012) and found that situations that involve social cues elicit an automatic social reaction. While these models focused on computers, the following models describe research specifically aimed at virtual characters.

Agency Assumption

The agency assumption states that the influence of human-driven avatars will always be higher than the influence of artificial entities represented by embodied conversational agents (Blascovich, et al., 2002). De Melo and Gratch (2015) describe agency as the ability to plan and act. In their research, agency is often established by researchers indicating to the participant that their interaction partner is a live person (has agency) or is a conversational agent (AI), even when that information is misleading. In the context of the agency assumption, social influence was assumed to be greater if the perceived agency, or mindfulness of a character was high (Blascovich, et al., 2002). Mind is represented through two core dimensions; agency and experience (de Melo & Gratch, 2015). Experience is described as the ability to sense and feel emotion (Gray, Gray, & Wegner, 2007). This research led to exploration into how to achieve believability and provide a virtual character with the illusion of life.

Believability – The Illusion of Life

Virtual characters are, quite often, meant to provide the illusion of life and support the viewer’s suspension of disbelief (or ability to believe the environment is real) (Bates, 1994). The goal is to apply strategies to convey a strong subjective sense of realism (Bates, 1994). Creating believable characters is something for which Disney animators have historically excelled. Disney's early animators (Thomas & Johnston, 1981) describe the emotionless character as lifeless and machine-like. The animator must consider these points: 1) The emotional state of the character must be clearly defined, 2) the thought process reveals the feeling and 3) The emotion must be accentuated and made apparent to the viewer immediately (Thomas & Johnston, 1981). These factors have been used to develop believable characters in animated movies, which can help give dimension to virtual characters. Meanwhile, Chuck Jones, animator of many Warner Brothers characters, including Bugs Bunny, argued that a flaw gives a character personality and that personality gives life (Bates, 1994). A good example of this is the character Miguel in the Pixar movie Coco (Disney Movies, 2017) who has a dimple on the left side, but not the right and whose smile is just a bit crooked. This imperfection improves the believability of the character. This is just one of many tools to make a character more believable. Aside from visual believability, there are other strategies to establish a believable character such as the appearance of reactivity; goals; emotions; and situated social competence (Bates, 1994).

deMelo and Gratch (2015) consider believability a construct that is difficult to define and measure. Instead, they compared interactions with virtual humans to interactions with humans with the intent of establishing an indirect measure of believability. This strategy led to the continued evolution of the Ethopoeia Concept.

Revised Ethopoeia Concept

In the revised Ethopoeia concept, the premise is that individuals are more likely to socially interact with virtual characters when behavioral realism is high (Nass & Moon, 2000). In this case, the response was assessed along a range with social interaction occurring more or less based on behavioral realism (von der Putten, Kramer, Gratch, & Kang, 2010). It is unimportant if the interactive character is an agent or an avatar, as long as the character displayed a high level of behavioral realism (Moon & Nass, 1996).

Threshold Model of Social Trust/Influence

The question of how much realism is enough has led to the development of the Threshold Model of Social Influence which sought to establish what level of certain social factors were necessary to reach the benchmark measure of people interacting with a Virtual Agent (VA) as they would a human (de Melo & Gratch, 2015). They proposed agency and experience as critical factors to meet the benchmark. von der Putten, Kramer, Gratch and Kang (2010) found that agency; the belief they were speaking with a person who could plan and act, had the greatest effect on self-disclosure independent of whether the character was actually an agent or an avatar; and found that removing behavioral realism or fidelity did not change those results. Further details about behavioral fidelity and agency is described next.

Behavioral Realism/Fidelity

When a virtual character is interactive and demonstrates situational appropriateness, they are considered to have high behavioral realism or fidelity (Blascovich J. , 2002). At the same time, if a character has a great level of visual fidelity, but does something that doesn't seem quite right, the viewer may experience the uncanny valley (Mori, 1970). This is a state that is believed to exist where technology approaches the likeness of humans, making individuals who witness the character feel uncomfortable (Mori, 1970).

Fidelity is the "level of realism that a simulation presents to the viewer" (Feinstein & Cannon, 2014). Improved fidelity is becoming more attainable in simulated environments due to improved processing capabilities and modeling tools. This may improve both believability and training outcomes. Summers (2012) suggested that if the fidelity of the simulation in relation to real world cues is not accurate the result can provide negative training transfer or negative training. One compelling argument for increased fidelity comes from Vice et al (2011). When providing visual cues in a virtual environment, subject matter experts may be able to perform a task, but may not be able to articulate exactly what cues are needed to support decision making. The decision loop that Warfighters rely on is the Observe, Orient, Decide, Act (OODA) loop (Grant, 2005), so there is a need for the M&S tool to provide sufficient detail to decipher appropriate cues, which in some cases requires higher fidelity graphics. One example of a simulation that requires higher fidelity is the US Marine Corps Combat Hunter training program which focuses on battlefield situational awareness and observation skills (Vice, et al., 2011).

The training goals of an M&S training tool should be the determining factor for fidelity of a simulation (Maxwell, Griffith, & Finkelstein, 2014). At the same time, fidelity can influence the extent to which users are able to suspend disbelief that the virtual environment is real and that what happens within it is meaningful with respect to the intended learning goals. Given this, and the importance of visual social cues in interpersonal interactions, fidelity might be an important factor to consider in meeting the Threshold Model of Social Trust.

Agency

Research by Nowak and Biocca (2003) describes agency as what distinguishes between a human (having agency) and a nonhuman (no agency). Perceived agency is the extent to which an individual sees him/herself or the avatar as being “in charge” of the virtual experience (Banks & Bowman, 2013). deMelo, Carnevale & Gratch (2014) suggest that agency is related to intentional acts (associated with humans) as opposed to random acts (associated with algorithms). To provide more context, Gray et al. (2007) described adult humans as high in perceived agency while animals and babies are low in perceived and actual agency. They propose that both agency and experience (the ability to sense and feel emotion) are the determining factors that create the sense of “mind” (Gray, Gray, & Wegner, 2007).

Behavioral/Gestures

Conversational gestures might function as an indicator of a conversational partner having exceeded the threshold of realism. Conversational hand gestures are "movements of the hands that co-occur with speech but do not appear to be consciously produced by the speaker (Jacobs & Garnham, 2006). For the purpose of this paper, distinct facial expressions will be considered in the same light. Hand gestures fall into four categories: representational gestures that can be iconic, metaphoric or spatial; botanic or rhythmic gestures; emblematic, which can replace words in language; and interactive gestures (Alibali, Kita, & Young, 2000). For the purpose of this paper, interactive gestures are of greatest interest.

Multiple researchers have explored how the number of gestures varies based on the visibility of the listener (Alibali, Heath, & Myers, 2001) , (Ozyurek, 2002), (Krauss, Dushay, Chen, & Rauscher, 1995) (Rime, 1982). Jacobs and Garnham (2006) designed research that strongly supports the primarily communicative function of gestures. Their research indicated that: "More gestures were produced when the listener appeared attentive than when the listener appeared inattentive." And that "speakers adapt their gesture usage to the perceived requirements of the listener." Their research demonstrated that gestures occur to benefit the listener. If gesturing is a tool used to bolster communication, it can provide insight into variations in communication strategies and communication partners as is described in this work.

Jones and Garnham (2006) explored two different strategies to measure frequency of gestures. One was gestures per minute of speech and another was gestures per 100 words. Each strategy correlated closely, but the rate per 100 words accommodates changes in speech rate. Given that, gestures per 100 words was used in this research.

Neuroscience as a Tool for Assessing Social Interactions/Game Theory

One primary strategy to measure social interactions is to measure a participant's response to certain social games while interacting with either an agent or an avatar. A participant's choices can, within the context of the game theory branch of experimental economics, provide insight into neural mechanics associated with decisions related to "trust, reciprocity, altruism, fairness, revenge, social punishment, social norm conformity, social learning, and competition." (Rilling & Sanfey, 2011, p. 23).

Classical game theory, such as the Nash Equilibria (Gintis, 2014) predicts that rational, self-interested players will use various strategies to optimize their performance. Through the Nash Equilibria, players might be expected to use backward induction, where they visualize the end and work backward to plan the moves that will bring greatest success (Gintis, 2014). In reality, players do not perform based on this premise. Research finds that players make decisions that are less selfish and strategic (Camerer, 2003). This suggests that game choices can provide insight into the participant's thoughts and feelings. Further, certain games lend themselves to interpretation of interpersonal interaction and are often used to provide insight into perceived agency (Nowak & Biocca, 2003). Some examples of social games that provide insight into social interactions follow.

The Ultimatum Game

The Ultimatum Game is a two-player game where a sum of money is divided. The "proposer" specifies a division of the money and the "responder" can accept or reject the offer. If the offer is rejected neither player receives anything, but if it is accepted the money is split as proposed (Guth, Schmittberger, & Schwarze, 1982). According to the Nash Equilibria, the responder should accept any offer even knowing that the Nash Equilibria dictates that the proposer should offer the smallest non-zero amount. In observed behavior, however, most offers are a 50/50 split and when the proposer offers less than 20% of the total amount the responder rejects the offer about half of the time (Guth, Schmittberger, & Schwarze, 1982). This can be interpreted as the acceptor being willing to lose money in order to inflict punishment on the proposer. Nowak, and Biocca (2003) showed that the same pattern was demonstrated with virtual characters. However, it would not make sense to punish a non-human, computer-controlled entity, which suggests that the virtual character had reached the threshold of realism which invoked the will for vengeance in the human acceptor.

The Trust Game

Similar to the reciprocal exchange model in the Ultimatum Game, the Trust Game is another tool to explore social decision making. In the Trust Game, an "investor" provides a select amount of money to a "trustee". The money is then multiplied and the trustee can provide some portion of it (or none at all) back to the investor. There is only one turn during the game (Berg, Dickhaut, & McCabe, 1995) though some researchers have expanded to allow for multiple turns. The Nash Equilibrium and game theory would suggest that the trustee would not honor the trust (minimizing payback), so the investor would minimize the investment, possibly to zero. However, again, research shows that investors do invest and trustees do return money to investors (Berg, Dickhaut, & McCabe, 1995). When virtual characters experience this investment/payback pattern it can establish another benchmark into individual responses to virtual characters and a sense that those characters have exceeded a threshold to perceived realism.

Prisoner's Dilemma

The Prisoner's Dilemma (PD) is a game that has been used for the past seventy years to study social decision making as it relates to trust (Poundstone, 1993). Two "prisoners" (sometimes more) from a gang are imprisoned in solitary confinement and are unable to communicate with one another. There is not sufficient evidence to convict both on the principal charge, but there is enough to convict them both on a lesser charge. Both prisoners are offered an opportunity to either betray the other by providing testimony against them, or to cooperate with the other by remaining silent (Poundstone, 1993). A matrix describing the prisoner's options is shown in Table 1.

Table 1 - Prisoner's Dilemma Outcome (Poundstone, 1993)

Prisoner A > **Prisoner B ∨**	**Prisoner A stays silent** **(cooperates)**	**Prisoner A betrays** **(defects)**
Prisoner B stays silent **(cooperates)**	Each serves 1 year	Prisoner A: goes free Prisoner B: 3 years
Prisoner B betrays **(defects)**	Prisoner A: 3 years Prisoner B: goes free	Each serves 2 years

There is no opportunity for retribution or reward outside the game and neither prisoner has any loyalty to the other. The Nash Equilibrium assumes that the two prisoners take a rational approach and they both betray one another. However, studies indicate that humans display a systemic bias toward cooperative behavior (Fehr & Fisxhbacker, 2003) (Oosterbeek, Sloof, & van de Kuilen, 2003) with mutual cooperation occurring about 50% of the time (Sanfey, 2007). Some researchers modify the game by iterating it over several sessions, adding consequences to the decisions (Singer, et al., 2006). PD has also been used to determine an individual's acceptance of a character as having agency.

Neuroeconomic Research and Measurement

Neuroeconomic research uses the games described above, to name a few, to encode what occurs in the brain during the process of social decision making (Sanfey, 2007). Some of the existing themes associated with this research include social reward, competition, cooperation, coordination; and strategic reasoning (Sanfey, 2007).

Social Reward

The social reward theme suggests that the brain responds to rewards. As such, brain measurements would be strongest in the striatum, a midbrain section of dopamine cells (Cromwell, Tremblay, & Schultz, 2013). This mechanism may improve choices through learning based on reward and punishment (Ascoli, 2014). Neuroimaging research shows that the striatum experiences increased activation when a person is engaged in a social decision to reciprocate or not, but there is a decrease in activation when there is unreciprocated cooperation (Rilling, Gutman, Pagnoni, & Berns, 2002). As described above, individuals might gain satisfaction from punishing defectors even if the player experiences some loss in the process. This was demonstrated with a Positron Emission Tomography (PET) study conducted during the trust game (De Quervain, Fishbacker, Treyer, & Schellhammer, 2004). Despite losing points themselves in the interchange players chose to punish partners activating the caudate located near the center of the brain beside the thalamus (De Quervain, Fishbacker, Treyer, & Schellhammer, 2004). The striatum was also activated when participants received money and when observing a donation to charity, especially when the donation was voluntary (Harbaugh, Mayr, & Burghart, 2007).

Emotions

Emotions play an important role in social decision making that has largely been ignored in previous research. Competition, cooperation and coordination is a theme that supports exploration of emotional processes and reward mechanisms (Sanfey, 2007). The importance of the ventromedial prefrontal cortex (VMPFC), the orbitofrontal cortex, and the anterior cingulate cortex in social decision making was experimentally demonstrated when patients who suffered damage to the VMPFC were impaired in performing a gambling task (Bechara, Damasio, Tranel, & Damasio, 2005). Individuals reacting to unfair offers in the Ultimatum Game displayed greater activation in the anterior insula (via Functional Magnetic Resonance Imaging (fMRI), which is an area associated with emotion. The area experienced greater activation as the unfair offer increased (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). Further, the anterior insula was more active when engaged with a human than with a computer partner and was a strong indicator of acceptance or rejection of the offer with significantly higher activation indicating rejection (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). This pattern was also displayed in the iterated PD game; with a stronger anterior insula response to unreciprocated cooperation (Singer, et al., 2006). This seems to correlate with the concept that this game relies on intense speculation or "mindreading" and a sense of fairness which would mainly apply to an individual with agency.

Strategic Reasoning: Theory of Mind

Theory of Mind (ToM) is a concept associated with how humans process the intentions and actions of others (Sanfey, 2007) by building a mental model of someone else's mind to make inferences or attributions about their mental state (Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004). The primary brain areas associated with this task include the medial prefrontal cortex and the anterior paracingulate cortex (Frith & Frith, 2003). McCabe, et al. (2001), explored the activation of these areas when interacting with a human or a random device such as a computer. fMRI data was collected to test the research question that the prefrontal cortex is involved in the theory-of-mind processing of cooperative tasks. Subjects participated in a two-person game with various payoffs depending on cooperation/noncooperation. Subjects consistently attempted cooperation when interacting with the human counterpart. Interactions with the computer were based on a known and probabilistic strategy. The prefrontal region was more active with human-to-human interaction. When participants played non-cooperatively with a human counterpart there was no significant difference in prefrontal cortex activity compared to human-to-computer interaction, confirming the importance of the prefrontal cortex in cooperative decisions. In fact, there may be other areas of the brain not traditionally considered part of ToM areas.

A study by Rilling et al. (2004) sought to determine if the areas associated with interpersonal decision making were consistent across two different social interaction games and if activation occurred in other areas of the brain. Activation was detected in the anterior paracingulate cortex and the posterior superior temporal sulcus (STS). While there was activation for both human and computer partners, there was greater response when the partner was human in both games. This suggests that this neural system might also be activated during reasoning about unobservable states during nonhuman interactions (Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004). Areas activated during these interactions included areas that had not previously been reported in ToM tasks, specifically: the posterior cingulate/precuneus, the mid STS, and an activation from the hypothalamus, midbrain and thalamus through the left hippocampus. The anterior cingulate, discussed previously, is densely connected to the limbic system structures such as the amygdala and hypothalamus and appears to be activated by emotional stimuli (Bullmore & Sporns, 2009). The posterior cingulate is located directly behind the anterior cingulate and appears to have a role in cognition and affect, possibly linked to autobiographical memories, and especially those that have an emotional quality. It may be part of the "default mode network" which is a group of brain structures related to daydreaming or recalling memories (Bullmore & Sporns, 2009). The posterior cingulate and hypothalamus activation in the Rilling, et al., study (2004) may have been the result of feedback from human partners. This activation was not present during interactions with the computer and further analysis indicated the subjects may have been processing previously encoded information on the partner's behavior and intent (Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004). One consideration in ToM studies is the potential for affective bias and individual differences.

Mixed Methods Research

Mixed Methods research integrates both qualitative (narrative or experiential) and quantitative (numerical) data (Wisdom, Cavaleri, Onwuegbuzie, & Green, 2012). Elements of the qualitative and quantitative data complement one another to provide a more complete picture of the research problem (Zang & Creswell, 2013). This research strategy capitalizes on the strengths of each individual strategy while minimizing each strategy's weaknesses (Halcomb & Hickman, 2015).

Example of study with Mixed Method Design

An example of a mixed methods study in the use of a VA was conducted by Yokotani, Takagi & Wakashima (2017). They used VAs that were compared to real experts in an interview setting. The research is based on the Threshold Model of Social Influence, which states that individuals are more likely to treat a VA in the same way as they would treat a human with social norms if the character displays high levels of agency and behavioral realism (Blascovich, et al., 2002). As stated previously, high agency means that the participant believes the character is controlled by a real human rather than a computer. As stated previously, behavioral realism is the degree to which the character acts like a real human, this includes: facial expressions, gaze, listening behaviors, and gestures. Some individuals actually had more rapport with a virtual character controlled by a human operator when they thought it was a computer than an actual person. For example, Yokotani, Takagi, & Wakashima (2017, p. 6) explored whether the audio-visual VA was outperforming the real expert during a comprehensive mental health interview. Interviews with VAs and real experts were compared by evaluating participants' perceived rapport, negative emotional expression, and self-disclosure of mental health symptoms. They hypothesized that a participant would self-disclose major symptoms more often and in more detail to the real expert than to the VA. In fact, despite the disadvantage the VA had in its ability to build rapport and express emotion, there were significantly more disclosures of sex-related information to the VA. The researchers posited that this might be due to the sense of anonymity in talking about private matters with a non-human entity. The VA diagnosed three false negative cases and five false positive cases of alcohol abuse and four false positives on the topic of eating disorders. There was evidence that female participants disclosed severe symptoms of a personal nature to the female VA, but not to the male real expert. This gender factor may have confounded the data. The qualitative findings indicated greater perceived emotional warmth toward the real expert.

The VA was perceived as having low agency and low behavioral realism, so participants did not believe the agent was real, which likely was the reason there was little warmth toward the VA. The qualitative findings indicated that participants self-disclosed eating disorder symptoms more to the VA than the real expert, but this may have been linked to the gender difference between the male live expert and the female VA. The real expert experienced more mood and anxiety disclosure than the virtual agent. It is possible that participants were aware of the camera on them during both interviews so their lack of anonymity may have affected the results. The mixed methods comparison of the quantitative and qualitative findings both support *and* deny the threshold model. Sex-related topics did see an increase in self-disclosure to the VA, but there could be other explanations, such as the gender of the agent, as seen above. The results of the higher level of disclosure with the live expert on topics of mood and anxiety may be related to the interviewer's nonverbal synchronization behaviors that made the participants feel they were immediately understood. This study had limitations in the areas of participant sampling with a small effect size and some participants were accessed using snowball sampling (reaching out to circles of other participants). The experimental condition was weak in that there was only one human expert and the determination of diagnoses is somewhat subjective. It is unclear what specific features of the VA had an effect on participants. Future research was suggested to explore various diverse VAs and real experts. Finally, there was no verification of diagnoses through a third party, nor detailed physical examinations to confirm diagnoses. This study encourages further research into factors that encourage participants to self-disclose to VAs and what factors about that VA enables participant trust.

Current Study Design

This mixed methods study will address the cognitive cues of social interactions to better understand what measures can be used to determine that a participant perceived their interaction partner as having agency. Survey data will explore user's perception of the social interaction and what factors, if any, drove them to disengage from the interaction. This is a concurrent, convergent, mixed-methods design. Qualitative and quantitative data are gathered at the same time, analyzed separately, and then the results are compared (Halcomb & Hickman, 2015). The quantitative data is intended to be primary and is supplemented or validated using the qualitative findings (Greswell & Creswell, 2018).

The following chapter will present the exploratory research methods that can be used to generate hypotheses to support future research in evaluating metrics to understand agency.

CHAPTER THREE: METHOD

The research questions described in CHAPTER TWO were addressed through the study design described in this chapter. The goal was to explore various measurement strategies to inform hypotheses for future research.

Participants

Twelve participants (7 males and 5 females) from a local population participated in the study, their mean age was 39.6 years and the standard deviation was 15.18 years. Participants were not compensated for participation. The target age of the sample population was military age, or eighteen years or older. Seventeen percentage of the participants were black and 83% were white. 8% of the participants had doctoral degrees, 25% had master's degrees, 17% had bachelor's degrees, 17% had either associates degrees or professional degrees and 33% had some college. 50% of the participants were single, 33% were married and 17% were divorced. 50% of the participants worked full-time, 17% were both student and military, 8% worked full-time with multiple jobs, 8% worked part-time, 8% were homemakers, and 8% were retired. 58% of the participants never played video games, 17% play frequently each week, 17% play rarely and 8% play one to two-times per month. 33% of the participants had been sexually assaulted in the past, while 67% had not. None of the participants had worked in the mental health profession. Demographics of the participants are shown in Table 2.

Table 2 - Demographics of Participants (N=12)

Topic	Demographic	Quantity	Percentage
Gender	Male	7	58
	Female	6	42
Age	18-29	5	42
	30-39	1	8
	40-49	1	8
	50-59	4	33
	60-69	1	8
Ethnicity	White	10	83
	Black	2	17
Educational Level	Some College	4	33
	Associates/Professional	2	17
	Bachelor's	2	17
	Master's	3	25
	Doctorate	1	8
Marital Status	Single	6	50
	Married	4	33
	Divorced	2	17
Employment	Full Time	6	50
	Student/Military	2	17
	Full Time/Multiple Jobs	1	8
	Part Time	1	8
	Homemaker	1	8
	Retired	1	8
Game Usage	Never Play	7	58
	Frequently Play per week	2	17
	Play Rarely	2	17
	1-2 times per month	1	8
Experienced Sexual Assault	Yes	4	33
	No	8	67
	Prefer not to say	0	0
In the Mental Health Field	Yes	0	0
	No	12	100
	Prefer not to say	0	0

Data Management

All data collected throughout this study is housed in a restricted-access location with password protection or keyed entry. Only those persons approved by the UCF Internal Review Board associated with this study have access. In addition, participants are assigned a unique identification number. No names or other forms of identification are stored with the data.

As part of the data collection and the analysis procedure, the participant is video recorded. The purpose of the video is to synchronize the information collected from the sensing devices with the participant's response to the interaction. The participant is able to request the opportunity to review the videotape once the session is completed. In addition, any video material will be erased after analysis is complete for this study.

Materials

Electroencephalography

EEG is the recording of electrical activity along the scalp produced by the firing of neurons within the brain (Abhang, Gawali, & Mehrotra, 2016). The research described in this paper will make use of the B-Alert X10 and the X24 EEG Headset Systems by Advanced Brain Monitoring. The system functions wirelessly through Bluetooth signal transmission and contains an 8-hour battery life. The sensors are lightweight and can acquire and analyze 10 and 24 channels of high-quality EEG data (Advanced Brain Monitoring, Inc, 2015). The sensor can be worn for over 24 hours and provides a comfortable and secure sensor-scalp interface during the study. The sensors are non-invasive and require no scalp abrasion. The system detects alpha, beta, delta, theta, gamma and high gamma frequencies (Advanced Brain Monitoring, Inc, 2015). Conductive gel is necessary for the sensors to record the brain signals. The gel used has similar ingredients as in human sweat and poses no more than normal risk as an allergen (Advanced Brain Monitoring, Inc, 2015). It can be washed out with water or can be wiped away with a paper towel. The EEG system has been tested and found to comply with the IEC 50501 safety standards (Advanced Brain Monitoring, Inc, 2015).

The EEG system requires that the participants undergo a 15-minute baseline procedure. The baseline tasks are presented via a computer and only require the participant to perform button press activity. Given that the tests measure user vigilance, the tests may cause drowsiness due to their repetitive nature. This drowsiness is not sustained nor expected to be maintained during the experiment (Advanced Brain Monitoring, Inc, 2015).

Neural Measures

As stated above, EEG measures electrical activity of the brain from electrodes on the scalp (Vogel, 2016). Pyramidal brain cells are perpendicular to the surface of the scalp and create electrical fields based on the firing of the neurons during response to various tasks which occurs in milliseconds (Rossini, et al., 2015). The electrical potentials produced by brain activity is amplified and output is created in a format for visual and data analysis as shown in Figure 1 (Niedermeyer & da Dilva, 2005). The system measures the activation of various areas of the brain at each of the range of frequencies. These activations are synchronized with the activities of the participant. These events are then compared for participants across each condition to explore the difference between conditions. These data will help to determine what brain events are activated in the various social interaction conditions of the study.

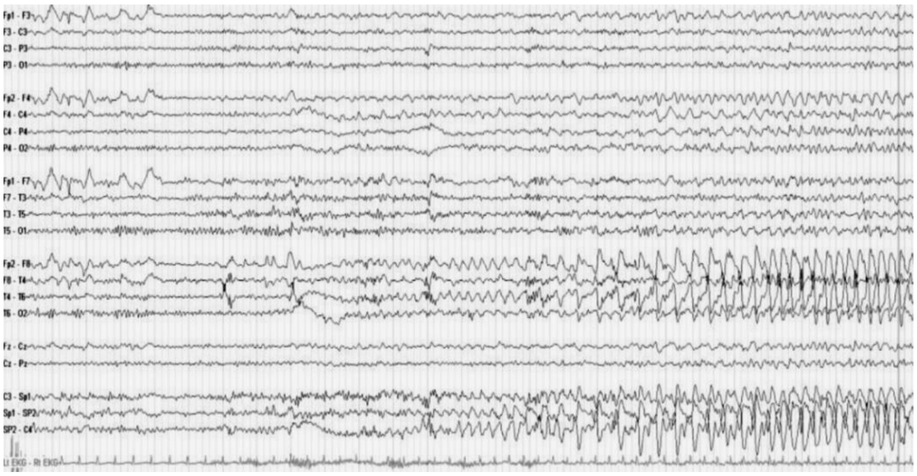

Figure 1 - Sample EEG Recording (Vogel, 2016)

Brainwaves

The electrical activity in the brain generates currents that flow with different frequencies and amplitudes. Each individual has their own natural flow of brainwaves which is why it is important to baseline an EEG system prior to collecting data (Abhang, Gawali, & Mehrotra, 2016). Electrodes pick up many waves with different characteristics. These are classified into gamma, beta, alpha, theta and delta. Figure 2 shows characteristics and waveforms of each wavelength.

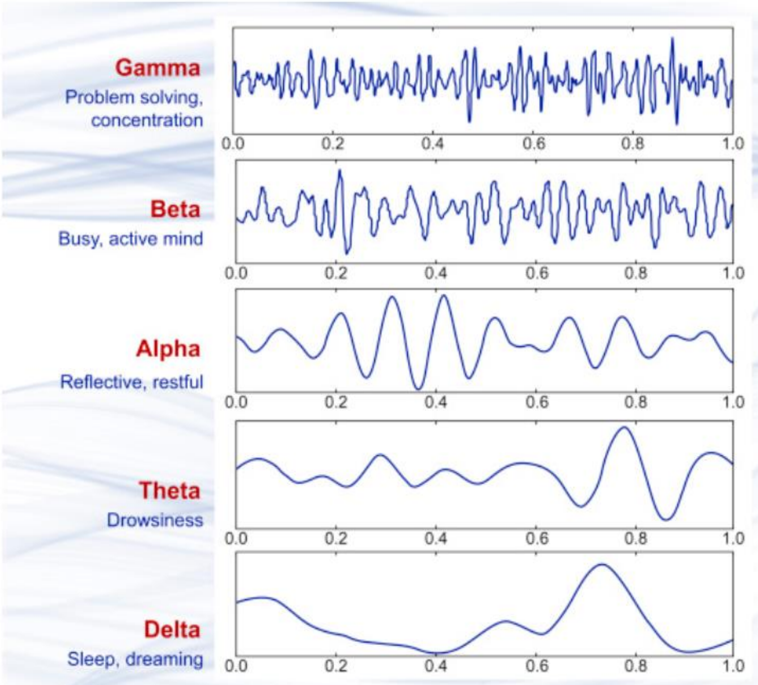

Figure 2 - Characteristics and Waveforms of Brain Wavelengths (Abhang, Gawali, & Mehrotra, 2016)

Gamma waves (25-100 Hertz) move fastest and are associated with coordinated processing of information from different brain areas. These waves are common during complex tasks, learning, information processing, problem solving, ideation, and concentration (Ismail, Hanif, Mohamed, Hamzah, & Rizman, 2016).

Beta waves (12-35 Hertz) are fast waves that indicates alertness. They are detected when an individual is engaged in problem solving or decision making. Beta waves are associated with logic, concentration, and productivity (Abhang, Gawali, & Mehrotra, 2016).

Alpha waves (8-12 Hertz) are often considered a bridge between the internal and external worlds (White & Richards, 2009). They are active when an individual is daydreaming, relaxing, visualizing or being creative (Ismail, Hanif, Mohamed, Hamzah, & Rizman, 2016).

Theta waves (3-8 Hertz) occur while sleeping but are also associated with deep relaxation, inward thoughts, memory recollection and emotions (Abhang, Gawali, & Mehrotra, 2016).

Delta waves (0.5-3 Hertz) often occur in deepest sleep but can also indicate fatigue (Abhang, Gawali, & Mehrotra, 2016).

The amplitude of these wavelengths increases with task load and may provide insight into what is going on in a participant's mind while they engage in an interaction. This information is collected through electrodes that are places along the skin on the skull.

Electrodes

Electrode placement is based on the International 10/20 system (Teplan, 2002). Each site has a letter to identify the lobe and a number to identify the hemisphere location with the numbers "10" and "20" indicating the distances between adjacent electrodes as either 10% or 20% of the total front-back or right-left distance of the skull (Trans Cranial Technologies, 2012). The lobe identifiers are shown in Table 3. It should be noted that the Central Node does not exist but is used for identification purposes. Electrode placement follows a precise numbering protocol. The "z" or zero electrode is placed on the mid-line. Even-numbers are on the right hemisphere, while odd-numbers are on the left hemisphere (Teplan, 2002). In some cases, notation may include a combined notation, such as Cp3 or C3'. This might stand for C3 prime or Central-Parietal 3, indicating that the electrode is midway between C3 and P3 (Vogel, 2016).

Table 3 - Electrode/Lobe Identifiers (Trans Cranial Technologies, 2012)

Electrode	Lobe
F	Frontal
Fp	Frontal Pole
T	Temporal
C	Central
P	Parietal
O	Occipital
A	Ear Lobe
Pg	Pharyngeal
Z	Zero (midline)

Electrodes are positioned based on four anatomical landmarks. The nasion is the point between the forehead and the nose; the inion is the lowest point of the skull on the back of the head and is characterized by a bump; and the pre-auricular points are directly behind the ear (Trans Cranial Technologies, 2012).

An example of sensor placement is shown in Figure 3. The number of sensors and their placement is dependent on the desired information of the study. Overlapping the functional areas of the brain and sensor locations indicate what sensors might be triggered based on specific behaviors. This is described further below.

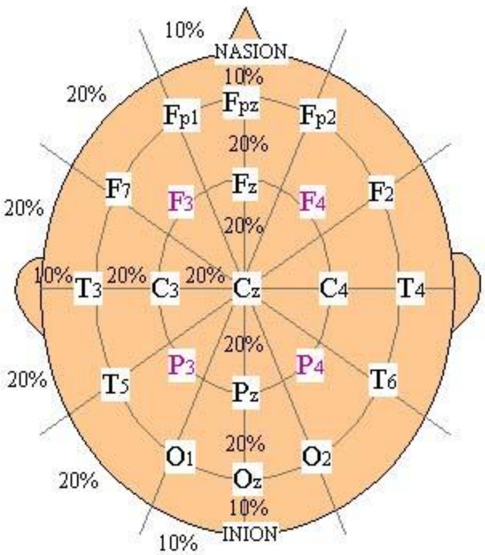

Figure 3 - Sensor locations based on the 10/20 standard (Teplan, 2002)

Since the occipital lobe, in the area of O1, Oz and O2 sensors, is associated with image recognition and perception we may see some differences when the participant is visually analyzing whether the person s/he is speaking to meets their threshold value of visual realism. The temporal lobe, in the areas of sensors T3 and T4, are associated with emotion, which would likely be activated during the interactions involving sexual harassment or assault in this study. The motor cortex (within the cerebral cortex) is covered with sensors C3 and C4 in the areas of the language center. The frontal lobe, in the areas of Fp1, Fpz and Fp2, is associated with higher mental functions, such as concentration, judgement and emotional expression (MidBrain Power, 2014). These areas could also be activated during the interactions in this experiment. Understanding the areas and function of the brain helps to ensure that there are sensors placed in the correct area to collect data to understand interactions between humans and various modes of interactions with virtual characters.

The B-Alert X25 applies more sensors along the frontal lobe, with fewer along the parietal area as compared to the B-Alert X10 (Advanced Brain Monitoring, Inc, 2015). The study will establish whether the additional sensors in the 25-channel B-Alert are a necessary as compared to the sensors included in the B-Alert X10 to capture differences between the study conditions. The participant used the X24 while the actor used the X10.

The plot in Figure 4 (A) shows summative 10-Hertz activity in the POz (middle scalp map) with the spectra of the projection to that channel of 32 lower traces and power maps from components 4, 5, 7 and 10. The plot in Figure 4 (B) shows the minimum and maximum values over all channels across time of five independent components contributing to the Event Related Potentials (ERPs) or the amplitude and latency measures of peaks in EEG measures (Delorme & Makeig, 2004). There are multiple analytical tools available to analyze EEG data including the open source EEGLAB analytical tools (Delorme & Makeig, 2004).

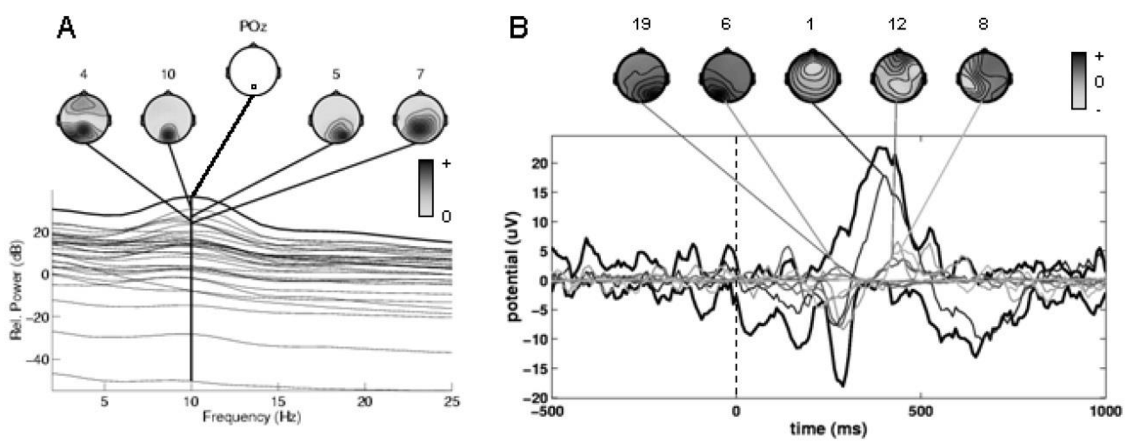

Figure 4 - Example of EEG data and the related area of brain activity (Delorme & Makeig, 2004)

Heart Rate and Inter-Beat Interval

Heart rate data was also collected by the EEG system. Heart rate variability measured ANS activity which is associated with stress (HRV Course, 2016). Heart rate variability was another physiological measure that was being assessed to determine if it was a reliable indicator that the interaction partner had exceeded the threshold level to convince the participant that they were interacting with a human being with agency.

Electrodermal Activity

EDA was measured using the Affectiva Q GSR Sensor. When "the sympathetic nervous system reacts, it causes many physiological changes including the release of miniscule amounts of sweat from sweat glands. These small changes of the skin's moisture change the skin and tissue conductance, which is measured by the sensor" (Neulog , 2018). The measure is based on quantity of moisture (sweat) on the skin from a continuous measure of the phasic aspect of electrodermal response (Hossain, Gedeon, & Sankaranarayana, 2016). The signal can be affected by minute changes in individual bodies, movement, sounds, etc. To reduce underlying noise from the GSR readings, a filter is applied (Hossain, Gedeon, & Sankaranarayana, 2016). The sensor measures the difference between the user's highest and lowest GSR values, commonly measured in micro Siemens (μS) (Noordzij, Scholten, & Laroy-Noordzij, 2012). GSR amplitude spikes have a rise time, amplitude, decay and amplitude shift from baseline (Neulog , 2018). The number and frequency of waves and amplitude shift were measured concurrently with the actor and participant.

EDA was measured by passing a small current through a pair of electrodes on the surface of the skin. Skin resistance is measured in voltage and is the reciprocal of skin conductance; this is expressed in units of micro Siemens (μS) (Dawson, Schell, & Filion, 2007). Skin Conductance Response (SCR) measures activation of the ANS, and in this case will be used as an indicator of emotional reactivity based on research that indicates the relationship between SCR and emotional arousal (Gregersen, Langkjaer, Heiselberg, & Wieland, 2017).

Surveys

Presence Questionnaire

The Presence Questionnaire (APPENDIX C) is a derivative of Witmer and Singer's (1998) Presence Questionnaire with items removed that are not applicable in this virtual scenario. The original questionnaire included factors across 6 subscales: involvement/control, naturalness, auditory stimulation, haptic response, resolution, and interface quality. Only four questions remained of this questionnaire:

1. How much did your experience seem consistent with your real-world experiences (naturalness)?

2. How involved were you in the experience (involvement/control)?

3. How much delay did you experience between your comments and expected responses (haptic response - however in this case it is intended to explore delays in verbal response)?

4. How well could you concentrate on the assigned task (interface quality)?

The questions were answered with a 7-point, Likert-type scale.

Rapport Questionnaire

The Rapport questionnaire (APPENDIX A) includes a set of questions associated with a sense of connection with another (von der Putten, Kramer, Gratch, & Kang, 2010). Each question is asked on a five-point Likert-type scale with 1 being "Strongly Disagree" and 5 being "Strongly Agree". Certain questions were reverse-coded (indicated in APPENDIX A with **R**). The average result of responses provided a relative number indicating each participants' perception of their rapport with their interaction partner.

Interaction Questionnaire

The Interaction Questionnaire (APPENDIX B) included both forced-choice and free-format open-ended questions that provide qualitative data to the study. Some of the forced-choice questions were collected from Artstein, et al. (2017), and include questions focused on the interaction partner's trustworthiness, persuasiveness, amiability, and the participant's enjoyment of the interchange. Other questions were more focused on the technology such as how effectively the participant was able to communicate and be understood. This questionnaire helped explore how aware participants were of emotional indicators, natural movements, and the naturalness of the actor's voice (though these were not applicable in the text communication condition). Frustration or stress associated with the interchange were also factors that were explored through this questionnaire. For each forced-choice question the participant was asked to provide open-ended responses to provide more insight into what they were aware of that affected their experience. The forced-choice questions were in the form of a seven-point Likert-scale asking about agreement with a statement with 1 being “Not at All” and 7 being “Completely.” Each question is distinct from the others. There was no attempt to establish inter-question reliability to create a survey product that could be used in future studies. Rather, this questionnaire was intended to dig deeper into participant sentiment to provide insight into what factors were meaningful in a communicative partnership.

At this point it is important to remind the reader of the participant's task. Participants were asked to conduct an interview for eight to ten minutes with Jarett, an individual who experienced a sexual assault. Participants were informed that they would need to relay what they learned and the general condition of Jarett to the proctor at the end of the interview. The format of the interchange allowed participants to control the dialog through the questions they chose to ask. This background might be useful to understand the responses to the open-ended portion of this questionnaire.

Social Phobia Questionnaire

The Social Phobia Inventory (SPIN) (APPENDIX D) is a validated survey (Connor, et al., 2000) that separates people into two categories, those with social phobia and those without. The survey contains 17 questions and was scored on a Likert-type scale with 0 being “Not at All” and 4 being “Extremely”. Questions include such items as “I am afraid of people in authority”. Summative scores greater than 19 indicated that the participant experiences social anxiety. This inventory is intended to tease out the role that social anxiety might play in the results of this study. Scores of participants with and without social anxiety were examined to determine the potential that this factor affected other findings.

Facilities, Proctor and Actor

Data collection occurred in an enclosed room with minimal distractions. A second room was used by the actor as appropriate. The actor was male and aged 50. One individual, the proctor, met with the participant and provided instructions.

Procedure

Tasks. Participants arrived and were provided a briefing on the study. They were warned that the subject matter included topics related to sexual abuse. The informed consent (APPENDIX G) was provided and briefed. Participants were reminded that they could abandon the experiment at any time and were asked to sign the informed consent form. Each participant was then asked to complete the Demographic Questionnaire (APPENDIX E), the Social Phobia Inventory (SPIN) (APPENDIX D) and the EEG Questionnaire (APPENDIX F) which requested data on the use of stimulants, such as coffee, sleep and medication that might affect EEG data.

The participant was fitted with an X25 B-Alert EEG headset then proceeded to baseline the system. The actor was fitted with the X10 B-Alert EEG and did not need to perform the baseline after completing it the first time. The baselining task took about twenty minutes. Participants were notified in advance of the expected time needed for this task.

The Affectiva Q GSR sensor was placed on the participant's wrist, then video cameras were turned on with one camera facing the participant and one facing their interaction partner.

The participant was told by the proctor that they would meet an individual who had experienced sexual abuse or assault. Their task was to ask the person for information about the event, giving the person the opportunity to share what happened and their feelings about it. They were told that they were gathering information because the participant would be answering questions about the event and the emotional state of the victim after the meeting. The intention was to motivate the participant to pay attention and engage the actor. The participant was able to choose the line of questioning, taking the conversation as deep or as safe as they chose. A timer was set to allow the discussion to continue for 8-10 minutes but was visible only to the proctor. This allowed the proctor to stop the discussions at a natural stopping point and avoid interrupting either speaker.

The particular interaction the participant experienced was randomly assigned. They experienced one of the below interactions:

- Face-to-face with a live interaction partner,
- Video teleconferencing with a live interaction partner,
- Interacting with an avatar that was controlled by a human puppeteer,
- Interacting with an algorithm-controlled character: DS2A application developed by the USC ICT,
- Or interacting via text with a live interaction partner.

The proctor started and stopped each session.

Layout

The layout of each interaction is represented in the images below (Figure 5-9).

Figure 5 - Participant Interacting with a Live Actor via Face-to-Face

Figure 6 - Participant Interacting with a Live Actor via Video Teleconferencing

Figure 7 - Participant Interacting with Human-Controlled Avatar

Figure 8 - Participant Interacting with Computer-Controlled Agent

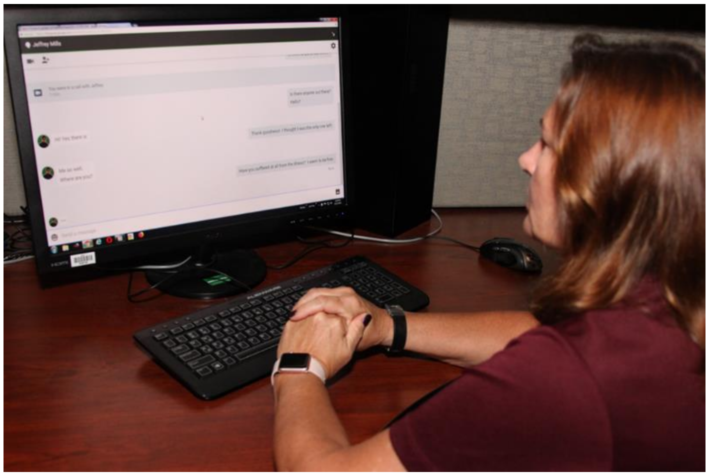

Figure 9 - Participant Interacting with Live Human via Text

Roles

Proctor – The proctor greeted the participant, informed them about the study and their rights for the informed consent.

Actor – The actor interacted with the participant in the various modes that required humans. He relayed "his" story of sexual assault. He wore an EEG and EDA sensors to determine if there was synchrony with the participant during the interactions.

Participant- This person wore an EEG and EDA sensors to measure social responses to the various interaction strategies.

Conditions

Face-to-Face Interaction:

The interactor and the participant sat comfortably between 2.5 to 3.5 feet apart at the knees. The actor introduced himself as "Jarett." Based on the questions asked by the participant, the actor described the sexual abuse scenario from the first-person perspective based on the story described by the DS2A application by ICT. After between 8 and 10 minutes, the proctor found a natural breaking point to avoid interrupting either speaker mid-sentence. The EEG and GSR sensors were removed. The proctor thanked the participant and provided the Rapport Questionnaire (APPENDIX A), the Interaction Survey (APPENDIX B) and the Presence Questionnaire (APPENDIX C) to be completed by the participant.

Video Teleconferencing:

The actor and the participant viewed one another through a 28-inch diagonal monitor sitting on a desk. The participant's seat was adjusted for comfort. The actor introduced himself as "Jarett." Based on the questions asked by the participant, the actor described the sexual abuse scenario from the first-person perspective based on the story described by the DS2A application by ICT. After between 8 and 10 minutes, the proctor found a natural breaking point to avoid interrupting either speaker mid-sentence. The EEG and GSR sensors were removed. The proctor thanked the participant and provided the Rapport Questionnaire (APPENDIX A), the Interaction Survey (APPENDIX B) and the Presence Questionnaire (APPENDIX C) to be completed by the participant. Avatar:

The actor controlled the virtual character and used it to interact with the participant. The interactor saw the participant through a video camera set below or above the monitor. The participant viewed the avatar through a 28-inch diagonal monitor sitting on a desk. The actor introduced himself as "Jarett." Based on the questions asked by the participant, the actor described the sexual abuse scenario from the first-person perspective based on the story described by the DS2A application by ICT. After between 8 and 10 minutes, the proctor found a natural breaking point to avoid interrupting either speaker mid-sentence. The EEG and GSR sensors were removed. The proctor thanked the participant and provided the Rapport Questionnaire (APPENDIX A), the Interaction Survey (APPENDIX B) and the Presence Questionnaire (APPENDIX C) to be completed by the participant. (APPENDIX A), the

Interaction Questionnaire (APPENDIX B) and the Presence Questionnaire (APPENDIX C) for the participant to complete.

Digital Survivor of Sexual Assault:

The participant interacted with the DS2A application developed by the USC ICT, through a 28-inch diagonal monitor sitting on a desk. The participant's seat was adjusted for comfort. In this case, the actual SFC Jarett Wright was the actual person the participants spoke to. An algorithm decoded the verbal questions and played the appropriate responsive recording for the participant. After between 8 and 10 minutes, the proctor found a natural breaking point to avoid interrupting either speaker mid-sentence. The EEG and GSR sensors were removed. The proctor thanked the participant and provided the Rapport Questionnaire (APPENDIX A), the Interaction Survey (APPENDIX B) and the Presence Questionnaire (APPENDIX C) to be completed by the participant.

Text:

The participant interacted via text with a human actor through a 28-inch diagonal monitor sitting on a desk. The participant's seat was adjusted for comfort. The actor introduced himself as "Jarett." Based on the questions asked by the participant, the actor described the sexual abuse scenario from the first-person perspective based on the story described by the DS2A application by ICT via text. After between 8 and 10 minutes, the proctor found a natural breaking point to avoid interrupting either speaker. The EEG and GSR sensors were removed. The proctor thanked the participant and provided the Rapport Questionnaire (APPENDIX A), the Interaction Survey (APPENDIX B) and the Presence Questionnaire (APPENDIX C) to be completed by the participant.

Wrap up

After the questionnaires were completed, the participant was provided a debriefing statement (APPENDIX H) to inform them that if they did not speak to the ICT DS2A then they actually spoke to an actor who relayed the information from DS2A to them as if it had happened to him.

Finally, it was possible that the study had invoked intense emotions within the participant so the proctor encouraged the participant to process their thoughts and emotions from the study. This was achieved by asking the participant to describe the scenario as they heard it from Jarett as well as any thoughts or emotions, they themselves experienced. This task provided an opportunity for the proctor to assess the mental state of the participant and determine if they needed more time to process or if they appeared settled. The proctor asked the participant about how they feel following the discussion and if they were alright. When the participant had fully discussed the experience, the participant was thanked for their time and informed that the session was over. If there was evidence the participant had been upset during the study, they were provided various resources to continue to process their feelings, such as contact information to UCF Counseling and Psychological Services (CAPS) and/or the Center for Research and Education in Sexual Trauma (CREST) in the UCF RESTORES lab. Additionally, resources such as the sexual assault hotline were on hand for participants who found that they were still negatively affected by the topic.

Risk

It was determined that there was no more than usual risk when wearing the EEG headset aside from some possible discomfort in wearing it. The risk of allergy could result from the gel, which was similar in content to human sweat. The risk estimate for allergic reaction to the EEG gel was determined to be less than 1% and all risks including discomfort were reversible. The overall risk was estimated at: Frequency: <1%, Severity: Minimal to no severity, Reversibility: >99%.

The psychophysiological sensors used in this study were determined to pose no more than minimal risk to the participant. The devices were disinfected and cleaned after each use. There has been no documentation of elevated risk associated with the use of these commercial devices. If any adverse situations had occurred, such as skin irritation, the incident would have been evaluated to ensure that the instance was isolated and that there was no damage to the equipment that may be the cause of such an issue. If the equipment had been found to be faulty, then a backup system would have been used.

CHAPTER FOUR: FINDINGS

Chapter Four presents the results of the various instruments, questionnaires and bio-sensors used to explore the research questions through quantitative methods and survey responses from the participants of this study.

Low Number of Participants

One issue with this study was the limited number of participants. Some factors at the root of this issue are that the Internal Review Board (IRB) process, between UCF and the Army took nearly 9 months. It took UCF the entire fall semester due to concern that the topic of sexual assault would cause distress to participants. The review board convened and conducted a question and answer period with the author, expressing concern for the wellbeing of people who might have experienced sexual assault themselves. These concerns were allayed by ensuring resources were available to participants if needed. Specifically, contact information to the UCF Counseling and Psychological Services (CAPS) and/or the Center for Research and Education in Sexual Trauma (CREST) in the UCF RESTORES lab were on hand. The RESTORES lab also agreed to provide counseling support to participants in need. Finally, resources, such as the sexual assault hotline, were on-hand for participants who found that they were negatively affected by the topic.

After UCF provided its IRB approval the U.S. Army Simulation, Training and Technology Center (STTC) that employs the author, was moved from the Army Research Laboratory under the Research, Development and Engineering Command to the Soldier Center under the Combat Capabilities Development Command (CCDC) under the U.S. Army Futures Command. The institutional agreement for IRB review was not yet established between the CCDC and UCF. Once this issue was resolved the entire spring semester had passed.

Another factor that influenced resources and personnel availability was that this was unfunded research. Participants were not paid and equipment was borrowed. In addition, four people; the actor, the participant, the proctor and the EEG technician had to be present at each session. Aside from the participant, each of these people had full-time jobs, health issues, travel demands, and family emergencies that made it challenging to schedule data collection. Once the sessions were scheduled, the set-up time for the 24-channel EEG was lengthy making it difficult to schedule more than two participants in a day. The chart in Figure 10 shows that in the six months that data collection took place no more than three participants were run in a month. When it became clear that data collection had reached the point of diminishing returns, the tough decision was made to proceed to analysis.

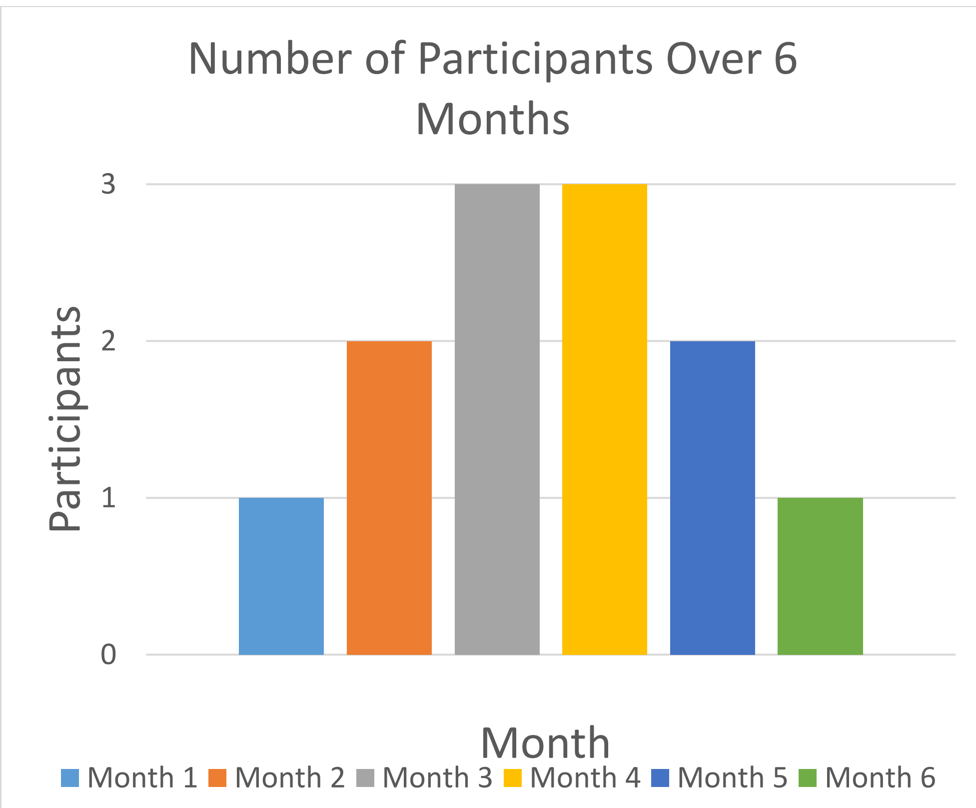

Figure 10 - Number of Individuals Participanting in Data Collection Over 6 Months

Each condition included either two or three participants with at least one male and one female in each condition. With so few participants, individual variation played too large of a role to determine meaningful conditional differences. The greatest challenges were in the IBI data. The interval for collection for analysis of IBI data was either one-minute or five-minutes. With this low number there was not enough data points within the 8-10-minute-long dialog interval to do an actual comparison between conditions. Despite these limitations, it was still possible to visualize the data and consider potential hypotheses for future research.

Electroencephalography

EEG Questionnaire

The EEG Questionnaire (APPENDIX F) was designed to consider factors aside from experimental conditions that might influence the EEG data. Factors include: caffeine, nicotine, alcohol, melatonin, marijuana, drugs, sleep, meals, stress, supplements, and which hand was the participant's dominant hand. Based on information provided by participants, estimates for quantities of caffeine, nicotine, and alcohol were calculated. Results indicate a range of caffeine from 0 mg to 166mg. The amount of nicotine ranged from 0 mg to 75mg. Melatonin was not a factor since none of the participants reported its use. Alcohol that was reported would have metabolized by the time of the study. One participant had marijuana in their bloodstream along with a variety of medications with unknown effects. One participant had taken Benadryl™ which might have a depressant effect and one had an amphetamine which could have a stimulating effect. Participants had between 4 hours and 10 hours of sleep the previous night. All but one considered themselves rested. Participants had finished eating between 1 and 16 hours before arriving. Each participant considered their stress level either medium or low. Some participants took a mix of vitamins, diet pills and fish oil with most participants not taking any supplements. Finally, there was only one participant who was left-handed.

EEG Results

The B-Alert X25 was applied to the participants and the B-Alert X10 was applied to the actor. The study data was used to determine if the X10 provided access to the social markers needed to assess cognitive social cues or if the X25 was necessary to generate meaningful data between the study conditions. The B-Alert EEG system detected cognitive state (sleep onset, distraction, low engagement, high engagement and drowsiness), brain wavelengths (delta, theta, alpha, beta, gamma and high gamma frequencies) as well as workload and heart rate data.

EEG is an individual measurement and could not be summarized for an entire categorized group. As such, each individual EEG result was included below in order of condition. The actor's data was included following each participant as appropriate (i.e. no actor data is included in the VA, or AI condition).

The 24-Channel EEG was cumbersome and time consuming to set up. Sometimes it took over an hour to get enough of the sensors to make sufficient contact. In one case this was not achieved during the first attempt and the participant had to return two months later, at which point the headset was eventually functional. This extended set-up time may have actually negatively affected the results, with participants experiencing irritation and fatigue. The 10-Channel EEG did not provide enough data across the frontal, temporal, and parietal lobes to collect sufficient emotion, attention, and workload information. Future research would likely benefit from a hybrid EEG system with sensors in optimized locations with the goal of reducing setup time while still collecting all pertinent data.

Brain wavelengths are shown as amplitude in microvolts (μV). Higher amplitude represents an increase in task load. The alpha and theta wavelengths were most common in this task due to alpha being related to visualization and theta being related to memory, emotions, and inward thoughts.

Following a long process, ranging between 30 minutes to over an hour, of getting the EEG headset in place, the baseline procedure takes another 15 minutes and is quite tedious. It involves three 5-minute sessions that require the participant to press keys based on different tasks. The first task involves recognition and response and is the most engaging. The next task is to press the space bar at a fixed interval based on a visual cue. The final task is to press the space bar at a fixed interval based on an audible cue with eyes closed. This might explain the level of fatigue found in several of the graphs that follow.

Condition 1 - Face-to-Face

Session 106

In the face-to-face condition the most likely cognitive state based on the probability value on the vertical axis was distraction as shown in Figure 11. It was also possible that the participant was fatigued which might have appeared as distraction. Probability values above 5.5 suggested that distraction was the cognitive state the participant was experiencing. This was further supported with the probabilities of other cognitive states being below .4 at their highest.

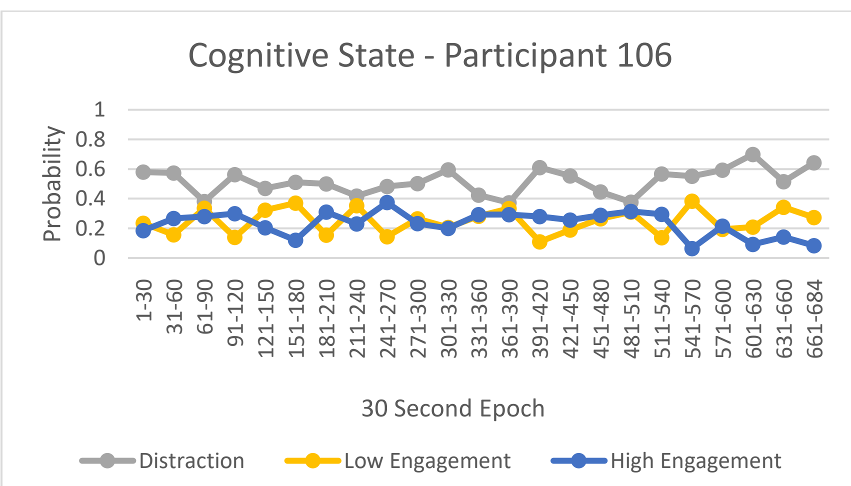

Figure 11 - Participant 106 Cognitive State

Participant 106 demonstrated predominantly alpha/theta brain wavelengths as shown in Figure 12. This combination suggested visualization combined with emotion or empathy. The participant's brainwaves showed signs (s)he was engaged.

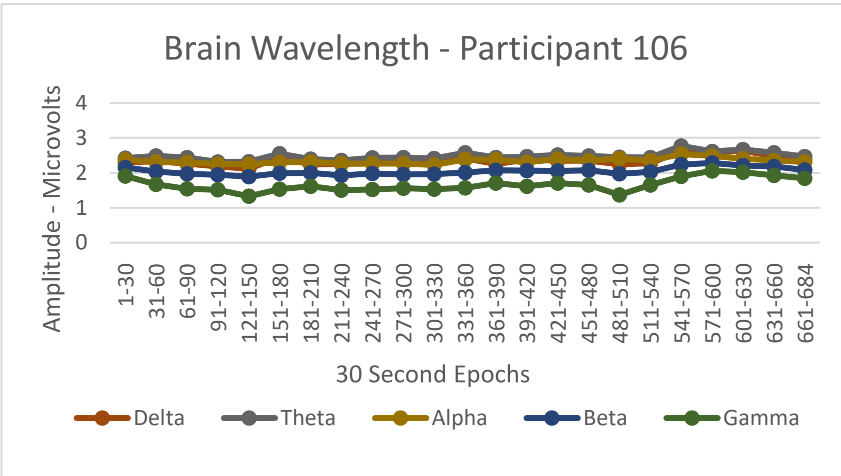

Figure 12 - Participant 106 Brain Wavelength

Workload above 5.5 was considered high. Figure 13 showed a moderately high level of workload for this participant as (s)he participated in this two-way dialog. There were areas that showed increased workload that correlated with the participant asking questions and experiencing awkward silences that drive him/her to speak, based on reviewing video footage of the interaction. This can be seen at the 541-570-time frame in Figure 13.

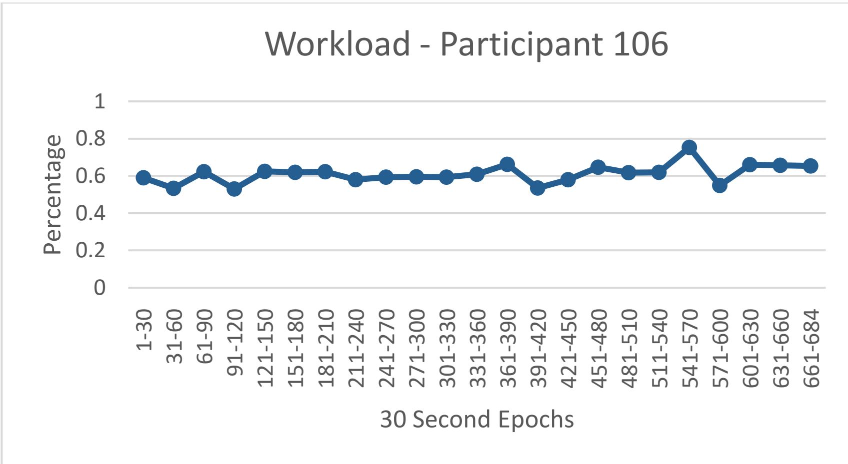

Figure 13 - Participant 106 Workload

EEG data for the same session was collected on the actor to look for similarities and evidence of synchrony. For a variety of reasons that will be discussed in CHAPTER FIVE, synchrony was not indicated during this study.

The actor's EEG data for the session follows. The cognitive state of the actor was fairly consistent from one session to the next (aside from a couple of notable differences) though they occurred on different days. In Figure 14 there was evidence that the actor was experiencing a high probability that low-engagement was his primary cognitive state. Distraction and high-engagement were both at low probabilities. This indicates that his level of engagement was moderate.

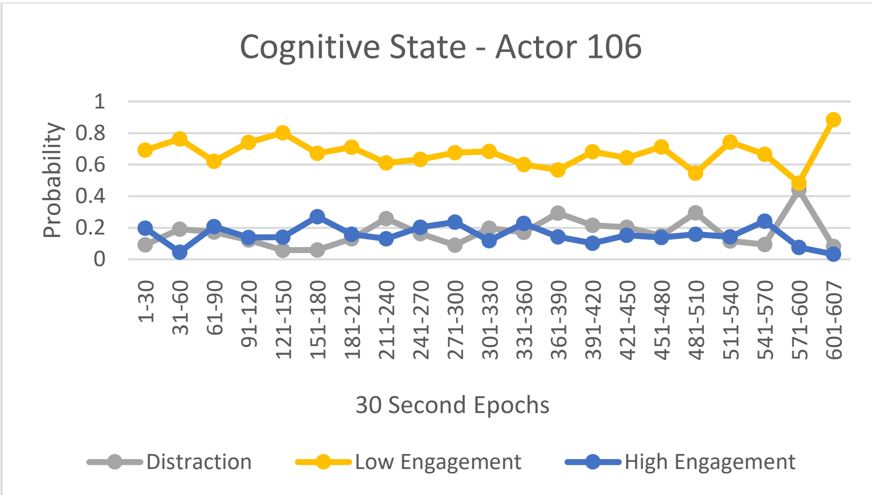

Figure 14 - Actor 106 Cognitive State

As described above, the brain wavelength shown in Figure 15 appeared to be the signature model for the actor in most cases. Gaps appeared in the data in two places and appeared to be related to noise in the system. In this case, alpha and beta had the greatest amplitude indicating greater task load in the areas of visualization, concentration, and alertness. This made sense as the actor had the task of applying his memory about the history and background of the real Jarett from ICT's DS2A to tell the story as if it happened to him.

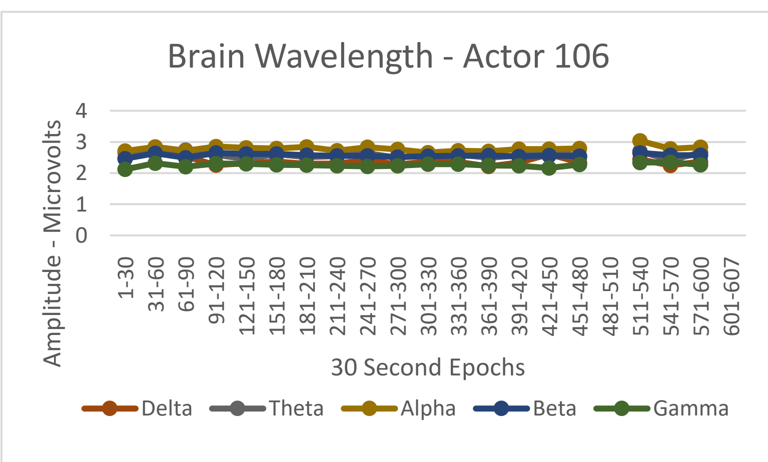

Figure 15 - Actor 106 Brain Wavelength

Figure 16 shows a consistent workload pattern on the part of the actor which can be seen in nearly all sessions involving the actor. His workload tended to be around 80 percent during these sessions. The actor produced, on average, an order of magnitude more conversational material than each participant for every condition except for the text condition.

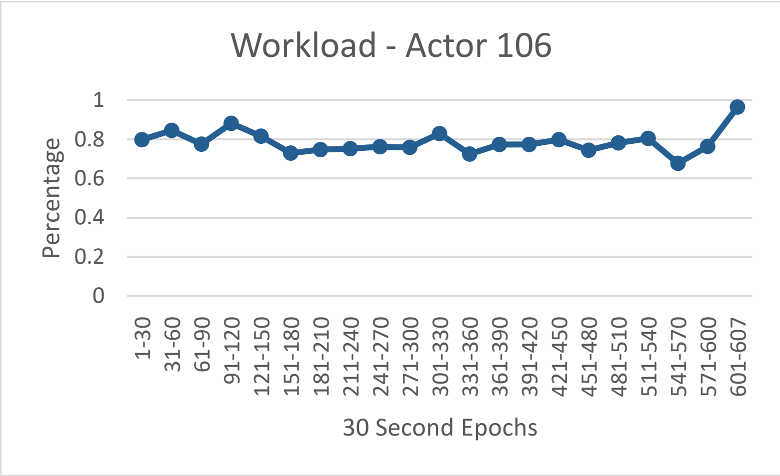

Figure 16 - Actor 106 Workload

Session 110

Participant 110 had a shortened graph as shown in Figure 17. Only about four minutes of data were captured as the EEG seemed to lose connectivity after that time passed. Despite that, it was clear that the participant moved back and forth between low and high engagement. The two lines were inverses of one another with low-engagement having the highest overall probability. This indicated a moderate level of engagement with the dialog. Two out of the three participants in this condition experience social phobia as was indicated in the Social Phobia Inventory (APPENDIX D) and this was one of them.

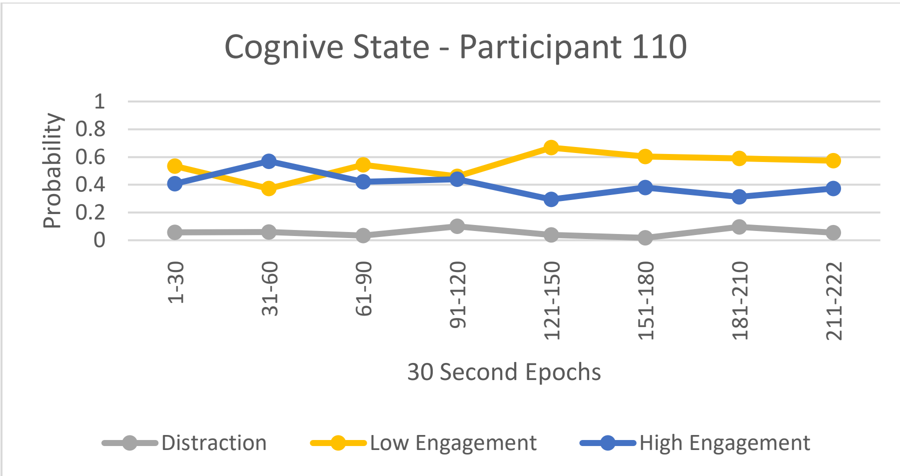

Figure 17 - Participant 110 Cognitive State

The primary activity in the brain wavelength graph in Figure 18 showed theta and alpha waves with alpha having a slightly higher amplitude at the beginning and theta having a higher amplitude at the end. Just after two minutes into the session, there appeared to be a separation of wavelengths with a jump in delta that might have indicated fatigue.

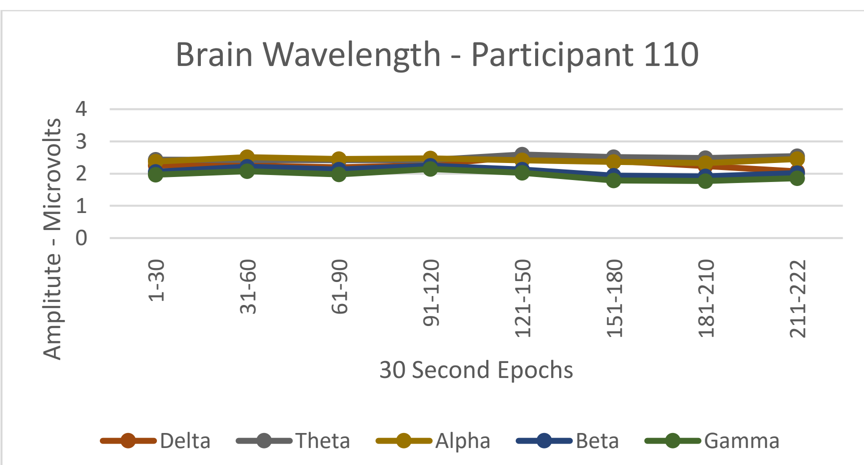

Figure 18 - Participant 110 Brain Wavelength

This participant's workload was moderately high as is shown in Figure 19. The peak in workload occurred at the same time that the separation in brain wavelength happened in the previous graph.

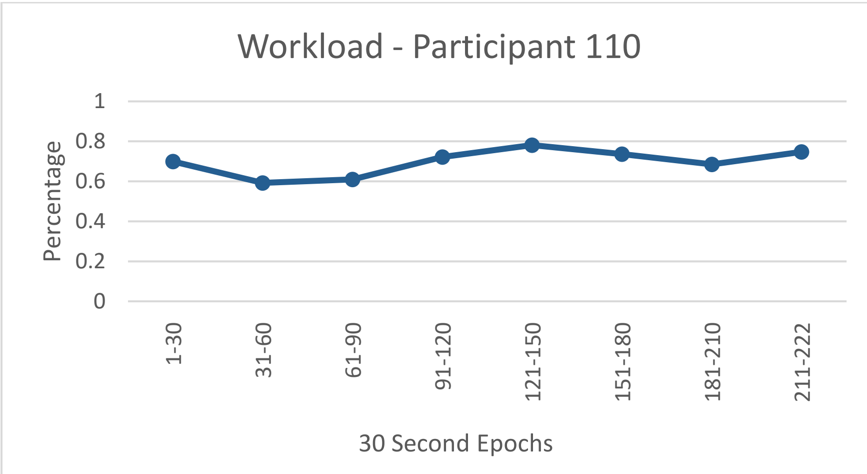

Figure 19 - Participant 110 Workload

The actor's EEG data for the same session follows. The cognitive state of the actor settled into his consistent pattern from that point on as is shown in Figure 20. Low engagement had a higher probability, than the other two states indicating a moderate level of engagement.

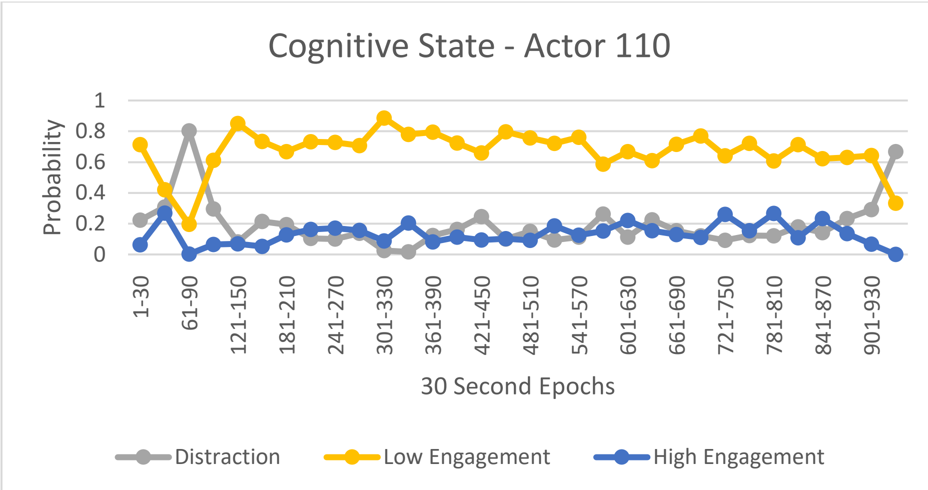

Figure 20 - Actor 110 Cognitive State

Gaps were shown in the brain wavelength patterns in Figure 21. This appeared to be the result of noise in the system, but aside from this noise, the pattern was consistent with the actor's patterns throughout the study. Alpha and beta wavelengths were present indicating concentration and visualization as he presented the story.

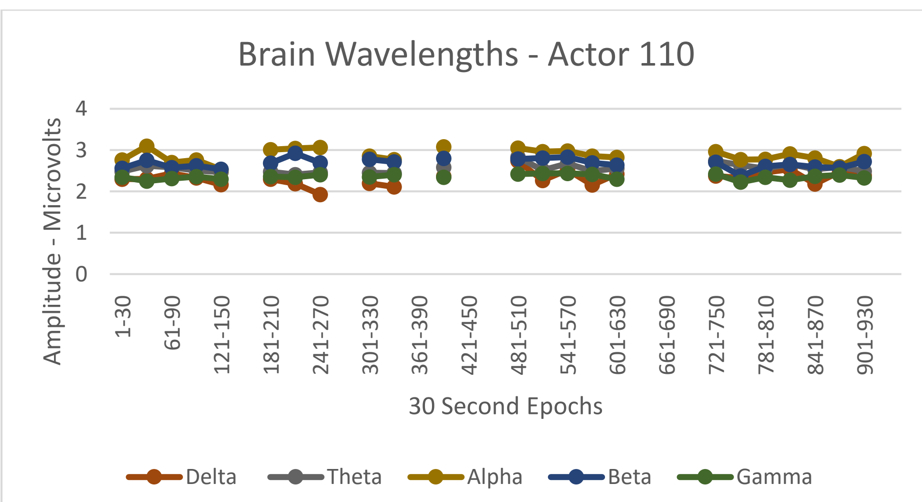

Figure 21 - Actor 110 Brain Wavelengths

Actor workload was high as shown in Figure 22 with peaks in workload being fairly close to the same peaks shown by the participant at the same interval. The actor's workload was generally high as he carried the greater share of the communication burden along with the need to remember and share the story.

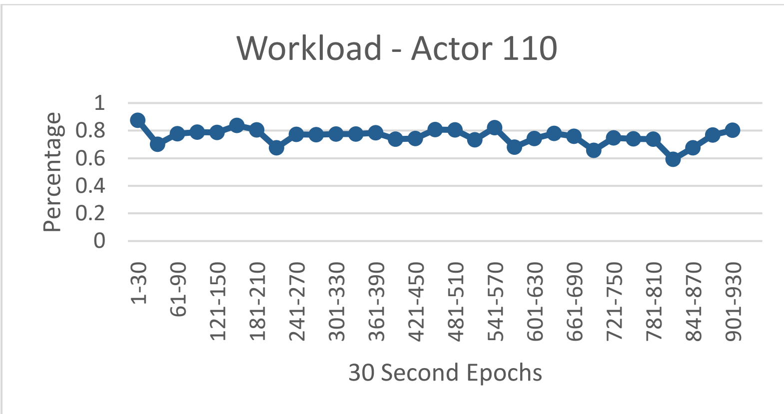

Figure 22 - Actor 110 Workload

Session 111

Participant 111 showed a high probability of distraction as shown in the cognitive state graph in Figure 23. The probability was very high that his/her level of engagement was relatively low and that (s)he was experiencing some level of distraction during the dialog.

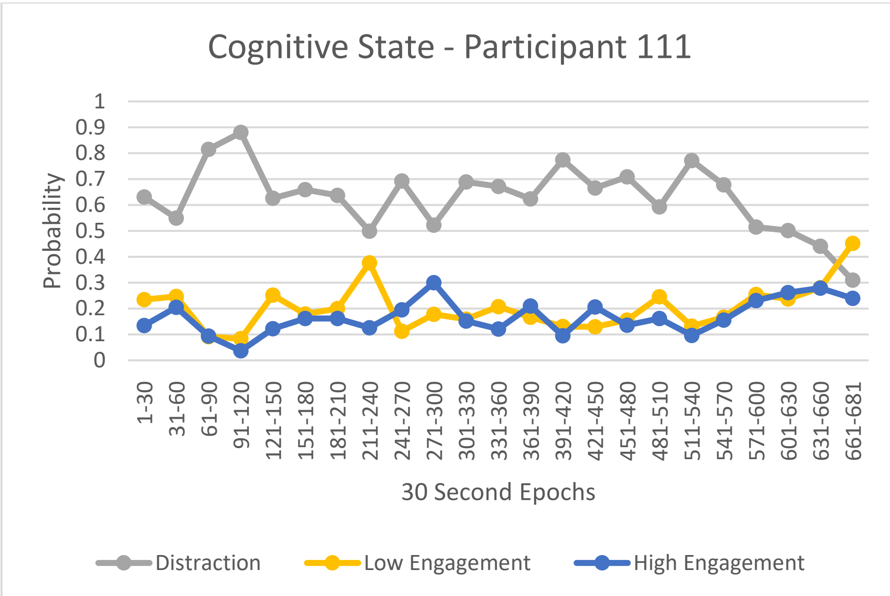

Figure 23 - Participant 111 Cognitive State

Alpha was the dominant wavelength as shown in the brain wavelength graph in Figure 24. The grouping of wavelengths was fairly close with little variation throughout the session. This participant had about 144 milligrams of caffeine in their system while participating in this dialog. Normally caffeine would be evident through in beta waves, but beta was not outstanding in this graph. Alpha and theta indicated empathy and emotion.

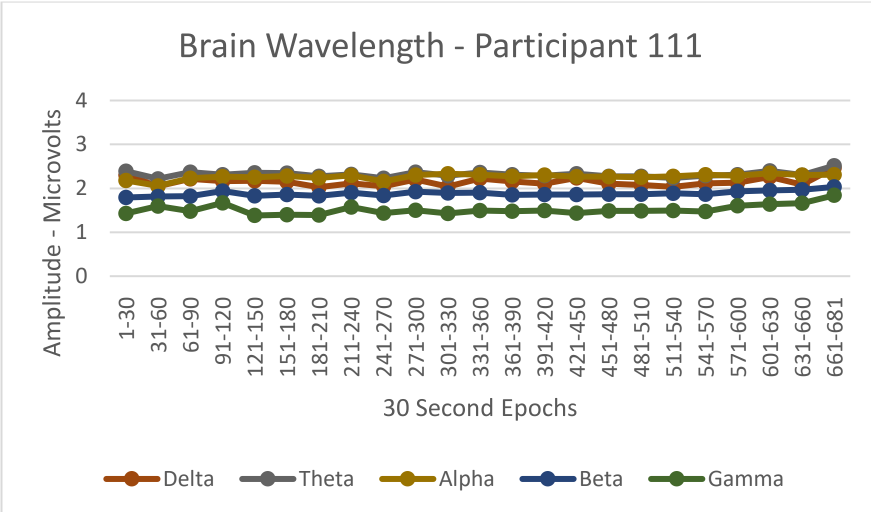

Figure 24 - Participant 111 Brain Wavelength

Workload was considered low since it was under .6 aside from a spike at the beginning of the dialog, as seen in Figure 25. This participant was taking a passive approach to the dialog.

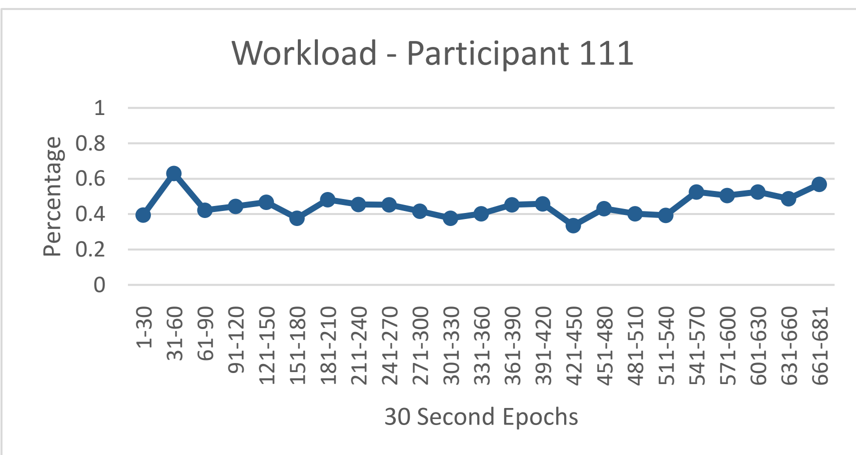

Figure 25 - Participant 111 Workload

The actor's EEG data for the same session follows. The cognitive state as shown in Figure 26 indicated that the level of engagement with the topic was in the medium range with a high probability of accuracy.

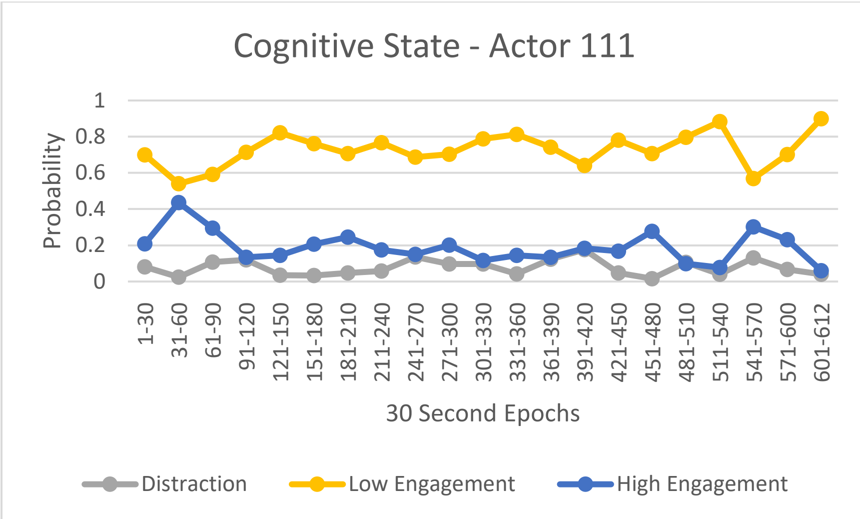

Figure 26 - Actor 111 Cognitive State

The actor was demonstrating alpha and beta at the highest amplitude in Figure 27. This supported the premise that the actor was concentrating on pulling remembered information from the DS2A application to respond to questions from the participant.

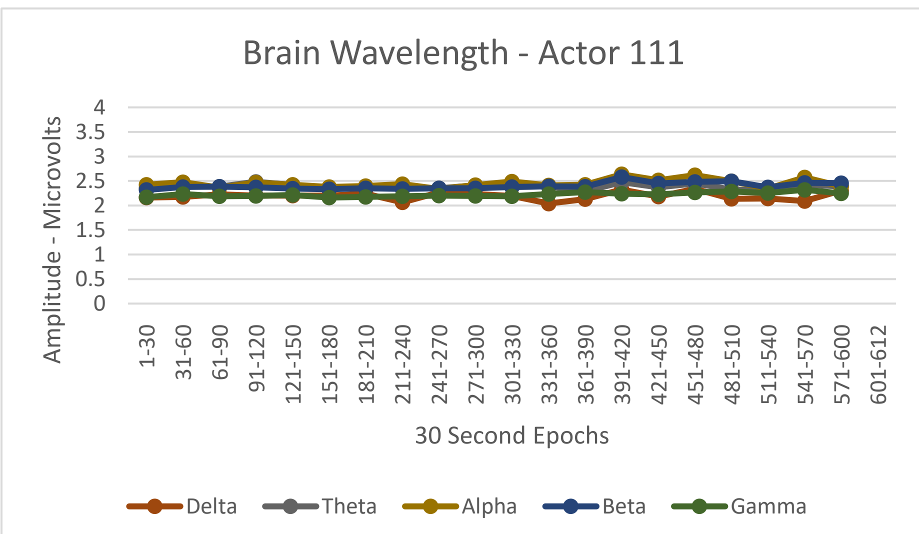

Figure 27 - Actor 111 Brain Wavelength

The actor's workload, shown in Figure 28, remained characteristically high for the actor, as he took the greatest burden of communication in the dialog.

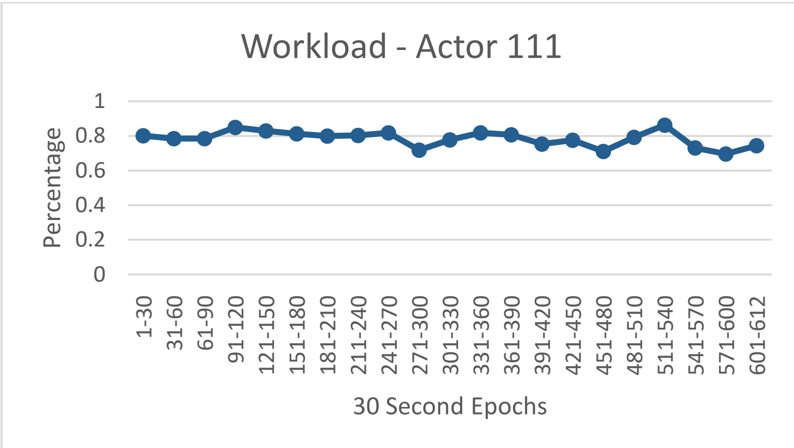

Figure 28 - Actor 111 Workload

Condition 2 – Video

Session 105

In the video condition, the first participant showed quite a bit of variation with high engagement having the greatest influence. There was a jump in distraction right around the middle of the interchange as could be seen in the cognitive state graph in Figure 29. The participant showed a high probability of high engagement throughout much of the dialog.

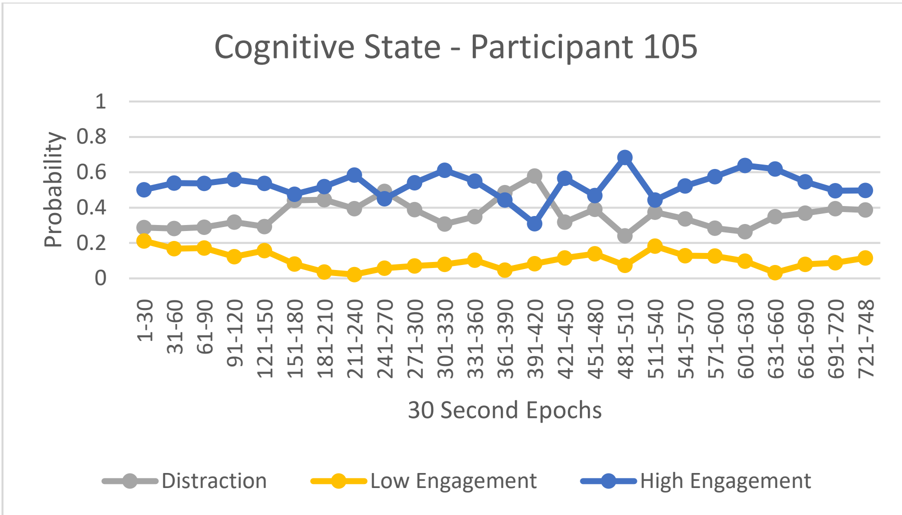

Figure 29 - Participant 105 Cognitive State

Alpha and theta had the highest amplitudes of the wavelengths. Delta also had a high amplitude suggesting a high level of empathy as was seen in the graph in Figure 30. It was possible that the participant was pondering the truthfulness of the actor during this session while balancing his/her empathy. This participant had about 72 milligrams of caffeine in his/her system during the interchange, but his/her overall beta was fairly low with minimal variation.

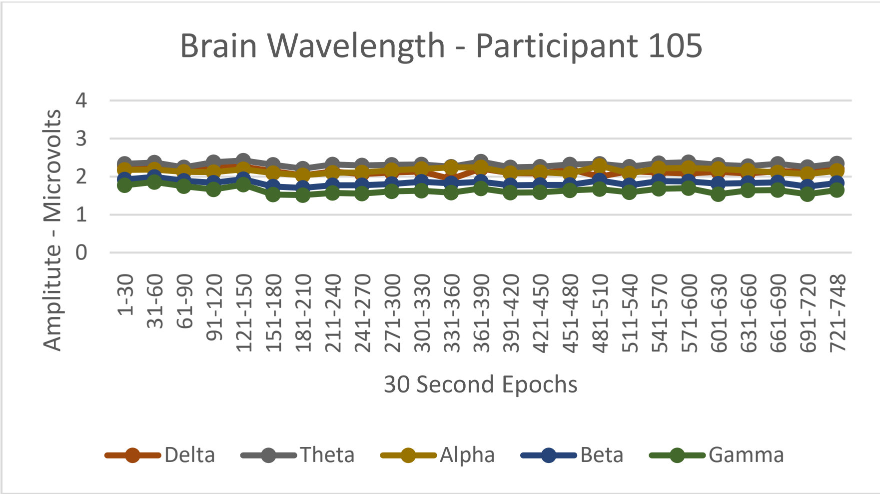

Figure 30 - Participant 105 Brain Wavelength

This participant's workload was on the high side of moderate workload as could be seen in the graph in Figure 31.

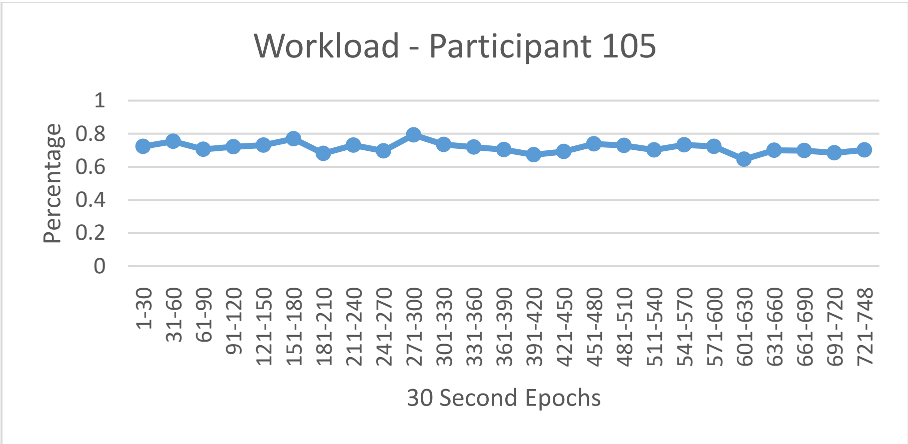

Figure 31 - Participant 105 Workload

The actor's EEG data for the same session follows. The cognitive state of the actor was consistent with previous sessions as shown in Figure 32. There was a high probability that the actor was moderately engaged with the dialog.

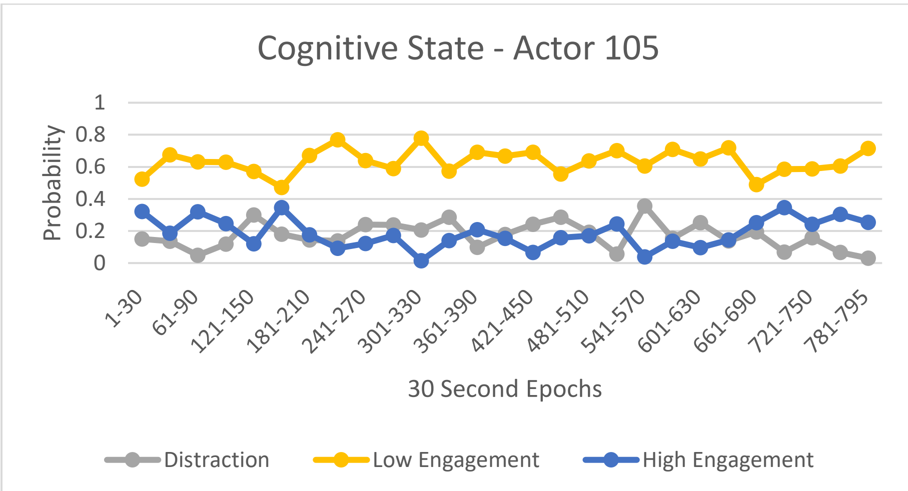

Figure 32 - Actor 105 Cognitive State

Similarly, the actor's brain wavelengths were consistent with other sessions with alpha, theta, and beta showing the highest task workload, as seen in Figure 33. These areas were consistent with concentration, memory recollection, visualization, and emotions.

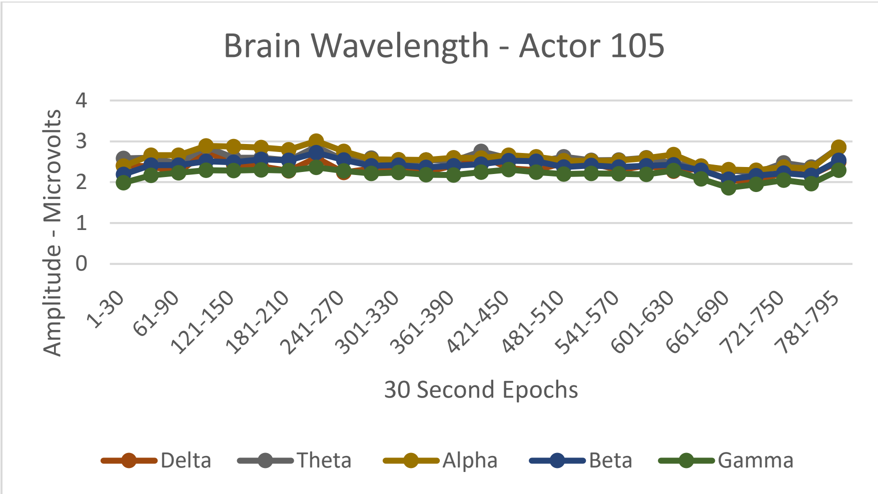

Figure 33 - Actor 105 Brain Wavelength

The actor's workload graph was consistently high, similar to other sessions as shown in the graph in Figure 34.

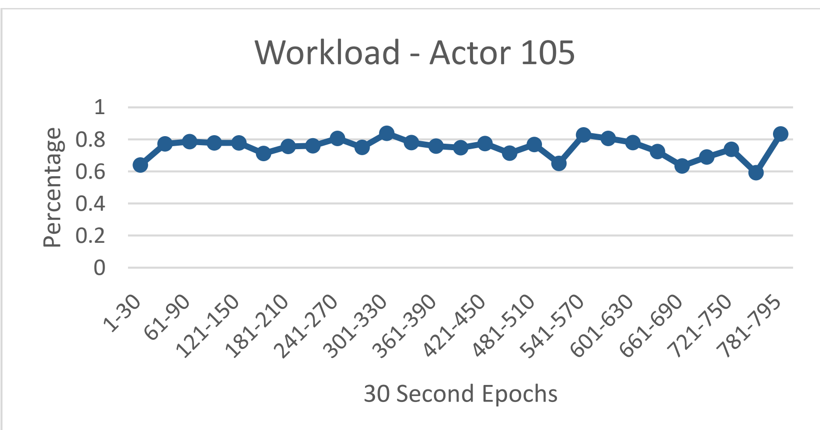

Figure 34 - Actor 105 Workload

Session 113

Participant 113 had a significant number of medications in his/her system that likely disrupted the EEG findings. The graph can be seen in Figure 35. (S)he was in the least engaged state. It was possible that fatigue played a role in this participant's experiences.

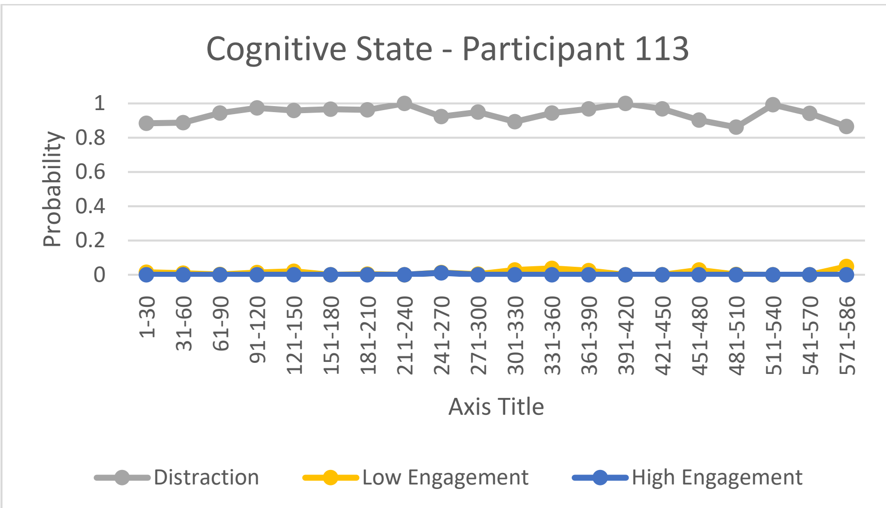

Figure 35 - Participant 113 Cognitive State

Alpha and theta were highest in the graph shown in Figure 36 with fluctuations in delta. As described above, there was a significant amount of medication influencing this participant's results, therefore it was difficult to put much meaning behind the graphs. However, the indications of empathy and emotional activation were still present.

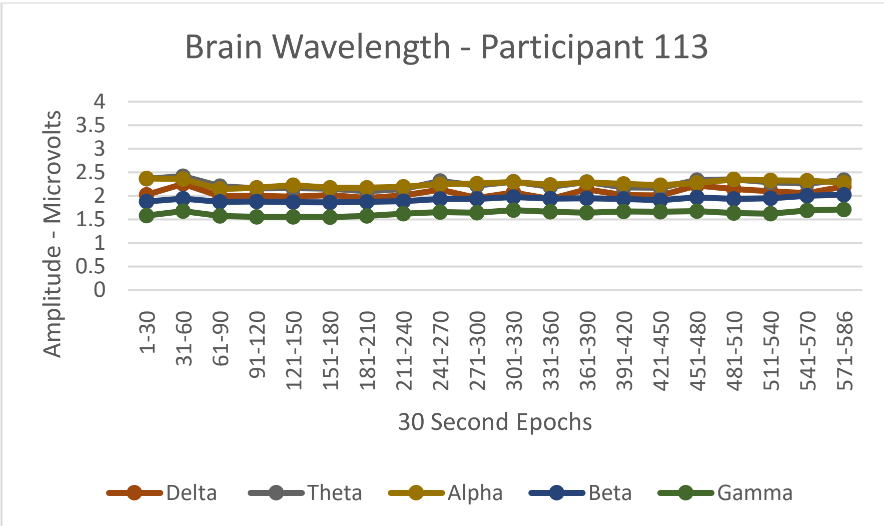

Figure 36 - Participant 113 Brain Wavelength

Workload for this participant was moderately high as shown in Figure 37.

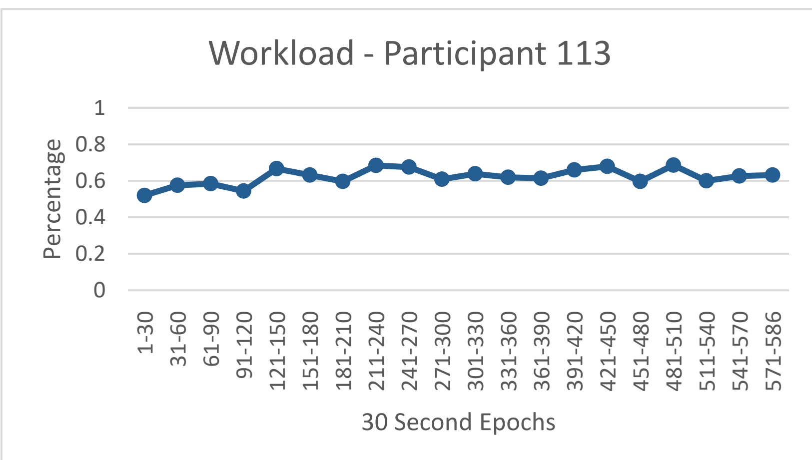

Figure 37 - Participant 113 Workload

The actor's EEG data for the same session follows. The cognitive state of the actor was not characteristic for him. It was possible that the cognitive state of the participant might have influenced the actor, as could be seen in Figure 38. At the same time, the actor did not complete the EEG questionnaire during each session, which means data were not available to determine if caffeine or medications might have affected his results.

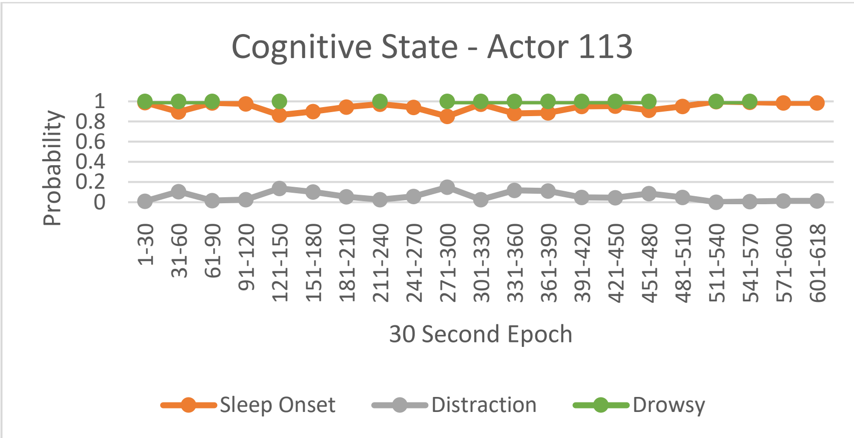

Figure 38 - Actor 113 Cognitive State

Similarly, the brain wavelength graph was not characteristic of the actor. Previous graphs of his brain wavelengths had been similar. In this case, delta and theta were tied closely together and took a primary position over the other wavelengths as seen in the graph in Figure 39. Delta was strongly associated with sleep, so it was possible that the actor was experiencing excessive fatigue on this particular day.

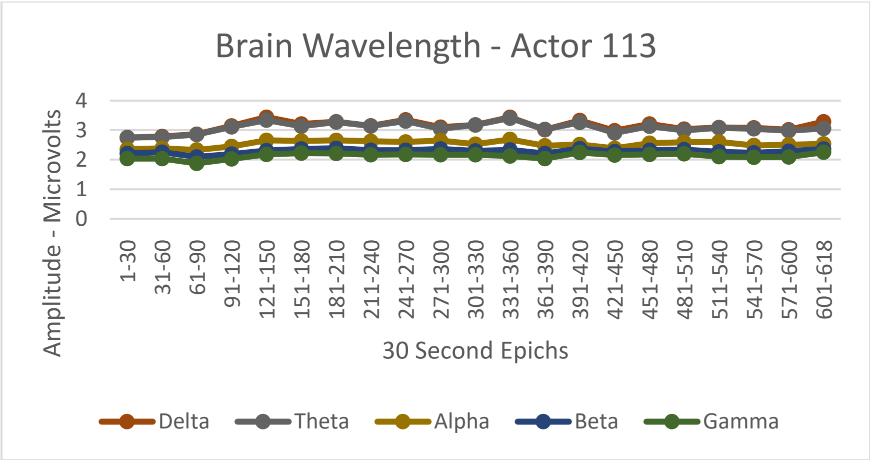

Figure 39 - Actor 113 Brain Wavelength

The actor's workload was even higher than usual. During this time, the participant was quite passively watching the interchange as if (s)he was watching a movie. His/her apparent engagement was so low that it might have influenced the actor's efforts to engage him/her as seen in Figure 40.

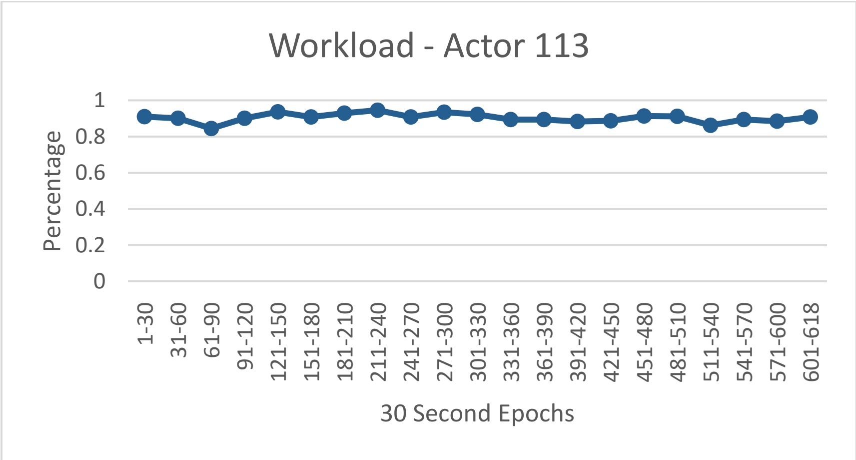

Figure 40 - Actor 113 Workload

Condition 3 – Human-Controlled Avatar

Session 107

In the human-controlled avatar condition, the first participant's cognitive state graph shown in Figure 41, indicated that (s)he had a high level of engagement, but that it went through peaks and valleys. Upon closer inspection the peaks of high engagement occurred when the participant spoke. (S)he was an animated speaker, possibly even impassioned at the greatest peaks when (s)he talked about how this should not have happened to the actor. The peaks in low engagement appeared to occur when the actor spoke for long periods. The probability readings were rather low so it appeared that the participant did vary in his/her level of engagement from high to medium.

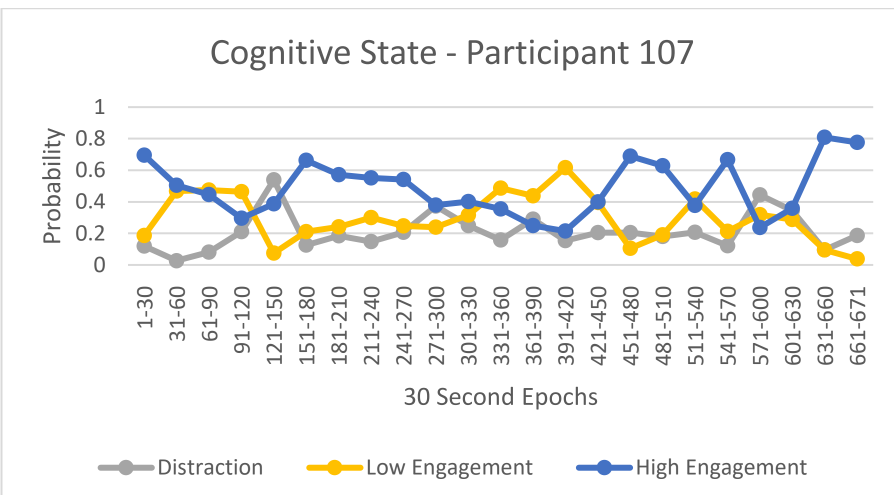

Figure 41 - Participant 107 Cognitive State

The brain wavelength activity shown in Figure 42 mirrored the behavior indicated in the cognitive state graph above. The wavelengths were tightly configured and moved in a consistent pattern with one another. Alpha and theta were primary again, similar to previous participants, suggesting empathy and emotion. Peaks and valleys in Figure 42 aligned with peaks and valleys in Figure 41.

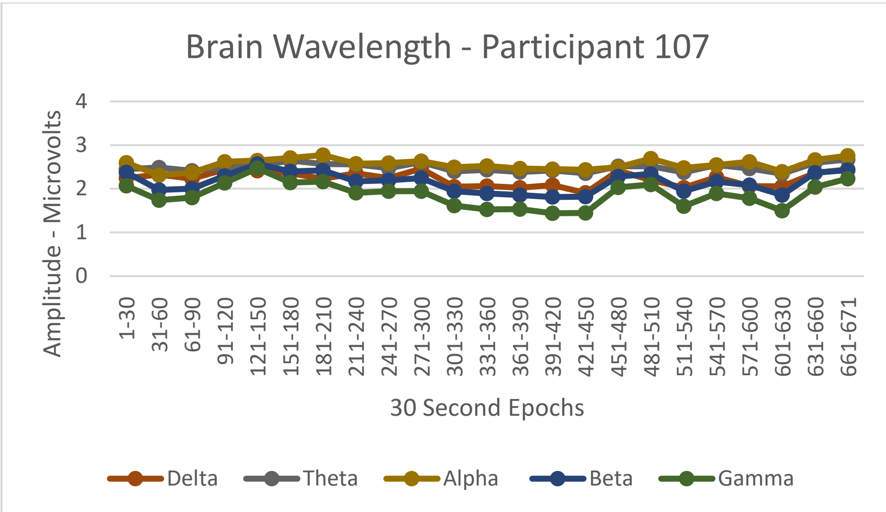

Figure 42 - Participant 107 Brain Wavelength

This participant's workload was high as shown in Figure 43. The peaks and valleys were mostly consistent with the previous two graphs except at around 331 to 360. During this time the participant wads helping the actor find the right word to finish a sentence. These charts indicated that this participant was fully engaged and it did not appear that speaking to an avatar was a detriment to that interchange.

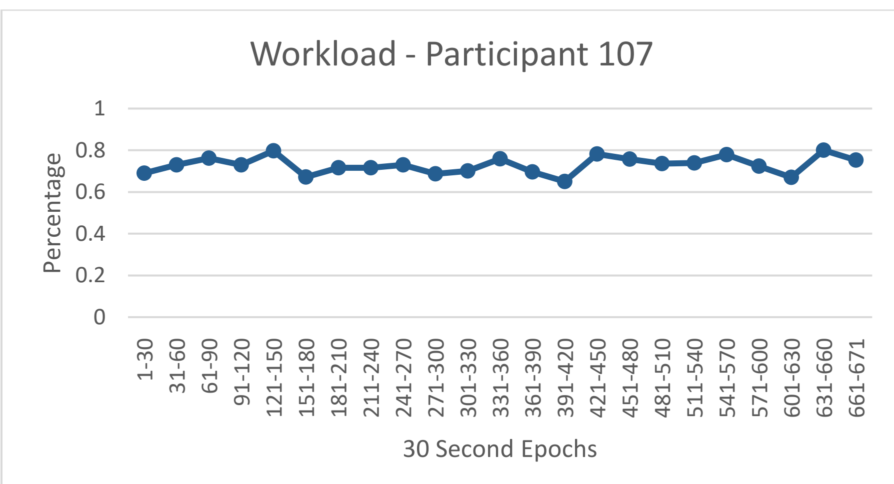

Figure 43 - Participant 107 Workload

The actor's EEG data for the same session follows. The cognitive state of the actor was not characteristic as shown in Figure 44. There was a high probability of low engagement with moments where the probability of distraction increased. When the actor takes over control of the avatar, he has the added burden of ensuring his arms do not twist in an unnatural way or cross into his body. This distraction from the interaction might have been what was seen in this graph.

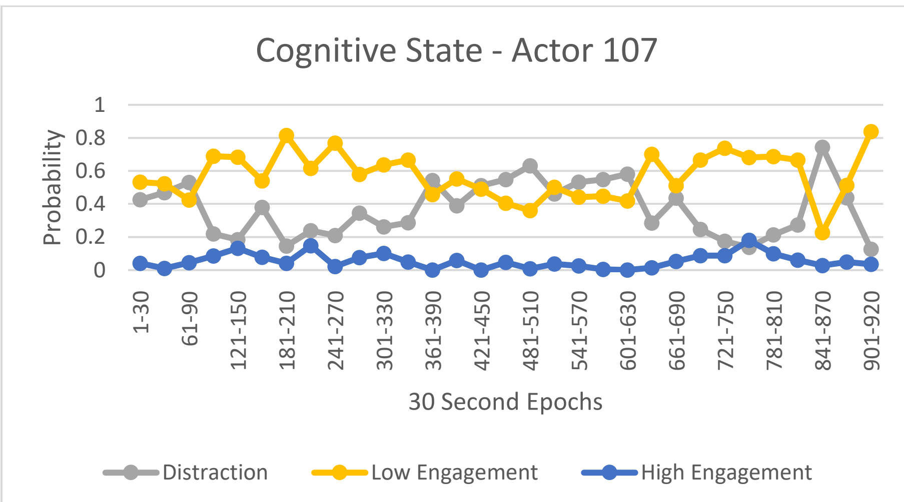

Figure 44 - Actor 107 Cognitive State

The actor's brain wavelength data was characteristic of his other sessions and is shown in Figure 45. His brain wavelength was tightly coupled and consistent with alpha and theta being primary, indicating empathy and emotion. It was possible that caffeine played a role in the actor's brain activity, but, as described above, caffeine consumption data were not collected for the actor.

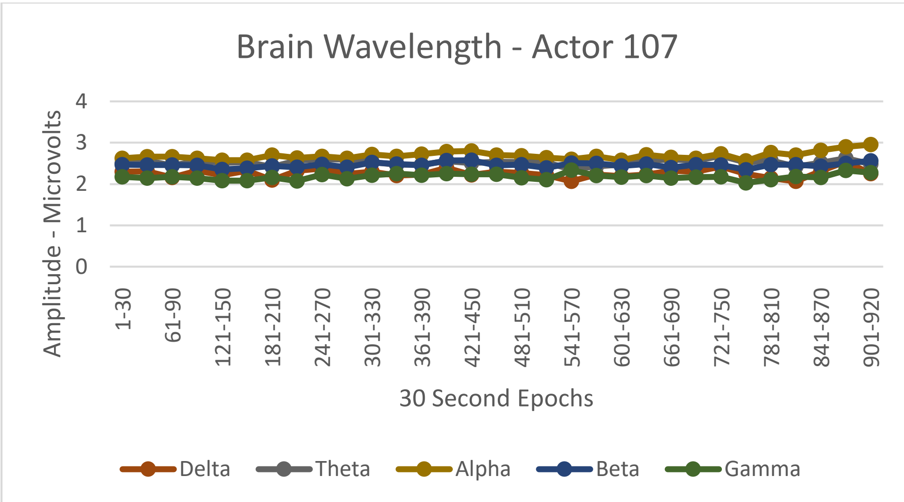

Figure 45 - Actor 107 Brain Wavelength

The workload data shown in Figure 46 showed significant peaks and valleys that the actor experienced. Even at its lowest, the actor's workload was high. Again, this likely had to do with the additional effort needed to maintain the dialog while ensuring the avatar character did not behave in an awkward way with arms colliding into the body or twisting unnaturally. It might be relevant to note that this was the first time the actor played this role while taking control of the avatar during this study.

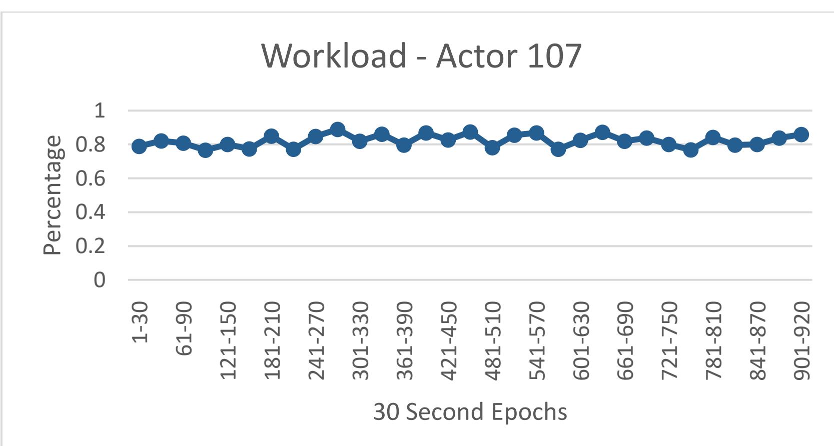

Figure 46 - Actor 107 Workload

Session 112

Participant 112 said the greatest number of words and displayed the greatest number of gestures of all the participants. His/her cognitive state, as seen in Figure 47, showed that when (s)he was speaking, his/her high engagement spiked and while (s)he was listening, low engagement rose. Interestingly, his/her gestures played a role in his/her communication, such as his/her eyebrows rising in a statement of shock. In some cases, his/her facial expressions said more than his/her words did. Since the probability was low that the cognitive state was accurate, it was likely that (s)he passed back and forth into different cognitive states.

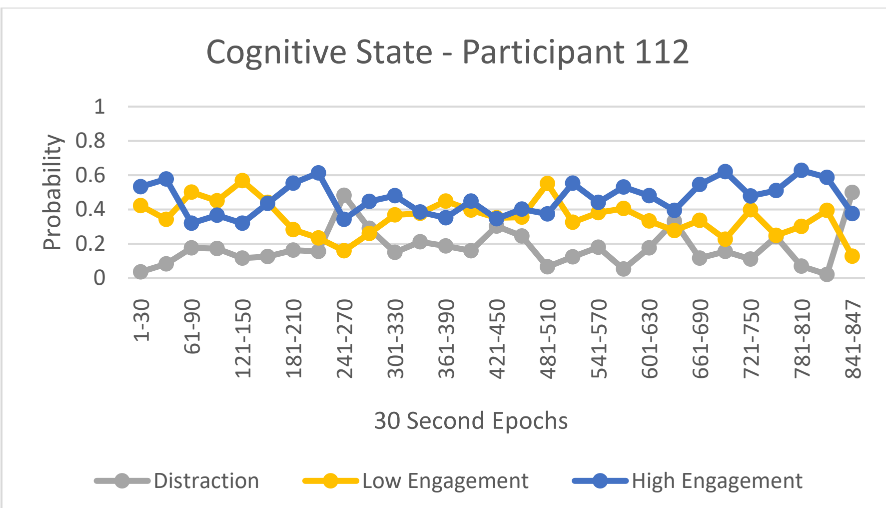

Figure 47 - Participant 112 Cognitive State

This participant's brain wavelength, shown in Figure 48, showed that theta, which was associated with memories and emotion was highest followed by alpha (visualization and problem-solving) and delta, which was closely related to sleep or fatigue. This was surprising since this person had 166 milligrams of caffeine in his/her system. It was interesting that gamma was the lowest amplitude wavelength on the chart, which would be related to information processing and ideation. The spike at 601-630 was at a turning point in the dialog where the participant expressed support and sympathy to the actor, bolstering his/her choices to speak out. Beta and gamma dropped at a time that the participant was speaking about the significance of male sexual assault in our society and how there was such a stigma associated with it.

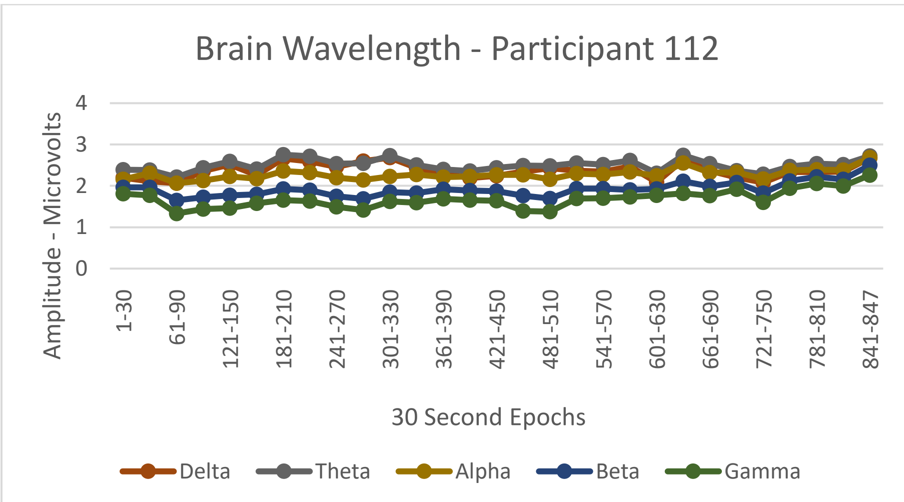

Figure 48 - Participant 112 Brain Wavelength

Workload for this participant was lowest in the second quartile of the discussion, after introductions and niceties passed and the actor told his story, as can be seen in Figure 49. The participant's workload increased as (s)he took a greater role in the dialog. It remained high suggesting engagement throughout the interchange.

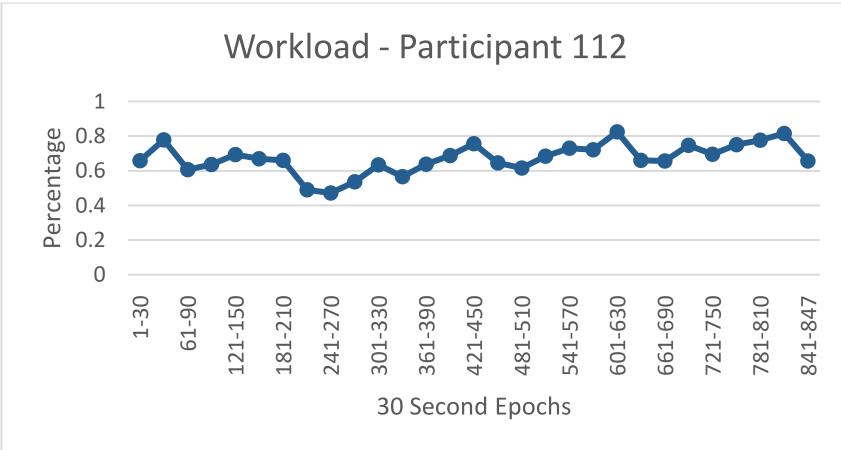

Figure 49 - Participant 112 Workload

The actor's EEG data for the same session follows. The cognitive state of the actor was back to his consistent pattern, with a high probability of low engagement being the mostly likely state as was shown in Figure 50.

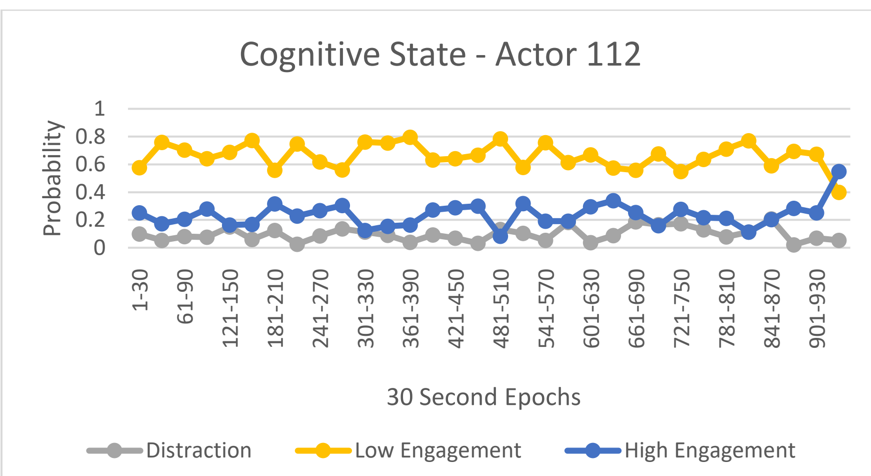

Figure 50 - Actor 112 Cognitive State

Brain wavelength for the actor was back to his standard profile with each wavelength being tightly coupled. Theta and alpha had the highest amplitudes as shown in Figure 51 suggesting empathy, emotion and memory recollection.

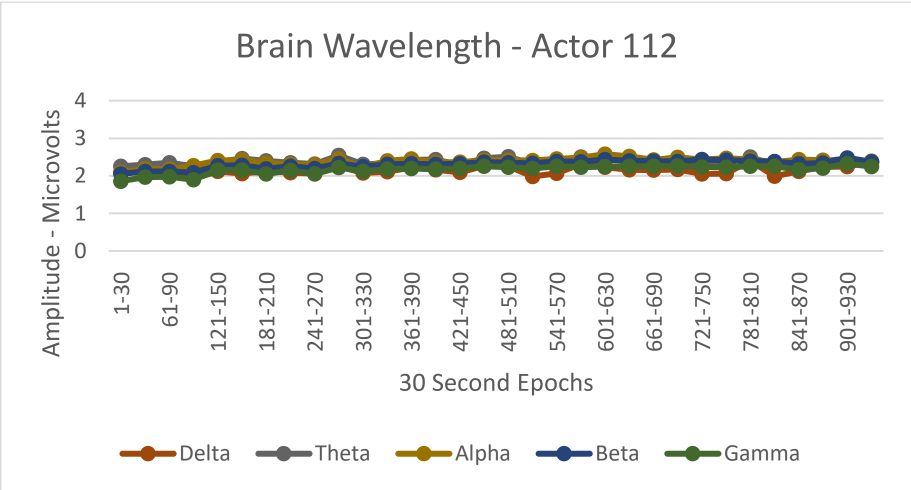

Figure 51 - Actor 112 Brain Wavelength

The actor's workload data in Figure 52 showed a consistent pattern with previous sessions. This session showed significant drops at the beginning and the end of the session. This coincided with waiting for the guidance to start and after wrapping up.

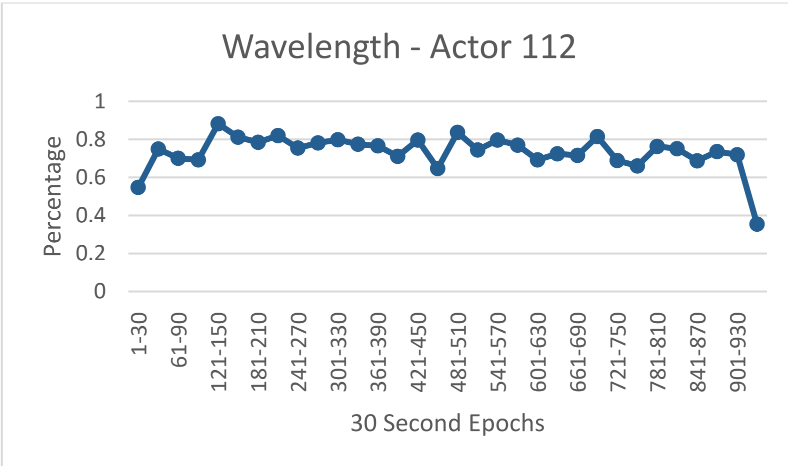

Figure 52 - Actor 112 Workload

Session 115

Participant 115 showed a high probability of the high engagement cognitive state as shown in Figure 53. At one point in the dialog, the participant had a moment where (s)he appeared fragile and withdrawn. That happened at the 361-390 timeframe. This participant had experienced sexual assault and it was possible that the dialog caused sensitivity.

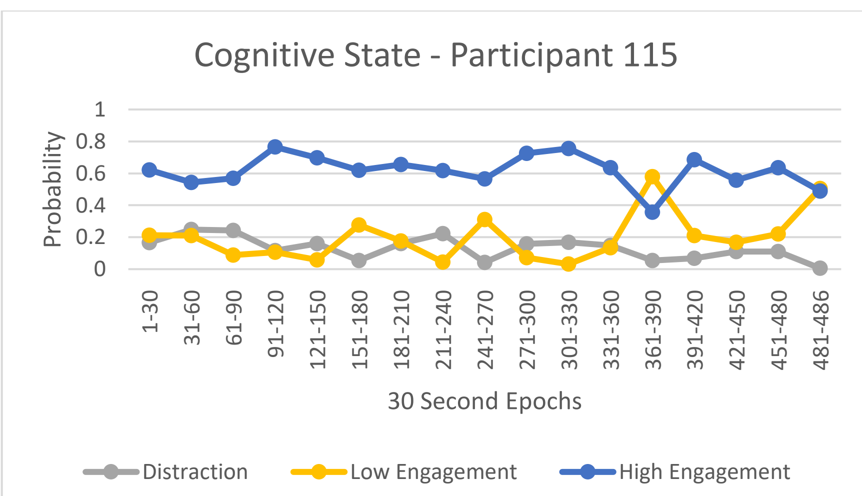

Figure 53 - Participant 115 Cognitive State

The brain wavelength graph in Figure 54 showed that alpha and theta had the highest amplitude, indicating empathy and emotion. This participant had roughly 104 milligrams of caffeine in his/her system, as well as 7.5 milligrams of nicotine. The caffeine was expected to show increased amplitude in beta wavelength, but that was not evident in this graph.

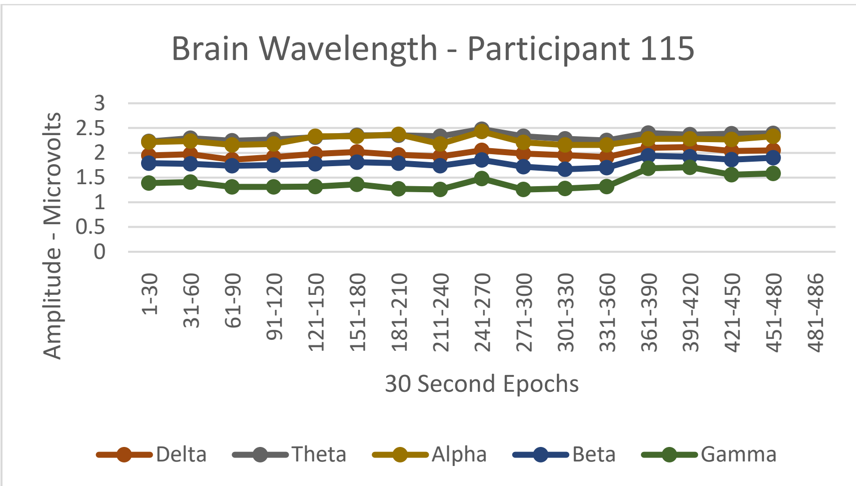

Figure 54 - Participant 115 Brain Wavelength

Workload for this participant was shown in Figure 55. Workload was high, indicating that (s)he was engaged in the conversation as an active participant.

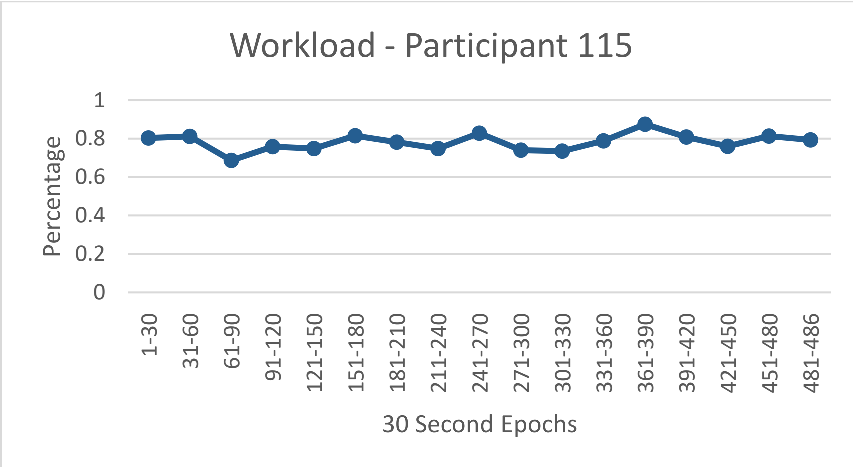

Figure 55 - Participant 115 Workload

The actor's EEG data for the same session follows. The cognitive state of the actor was consistent with previous graphs. This was the third time he has taken over the avatar, so it was possible that it had become more natural at this point. His data is shown in Figure 56.

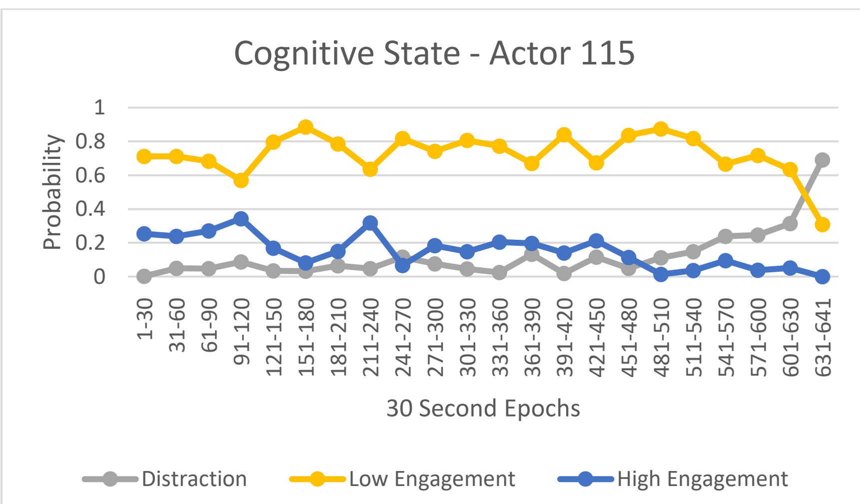

Figure 56 - Actor 115 Cognitive State

This was a common brain wavelength graph pattern for the actor. Alpha and beta had the highest amplitudes as he concentrated on both the story and controlling the avatar as could be seen in the results in Figure 57.

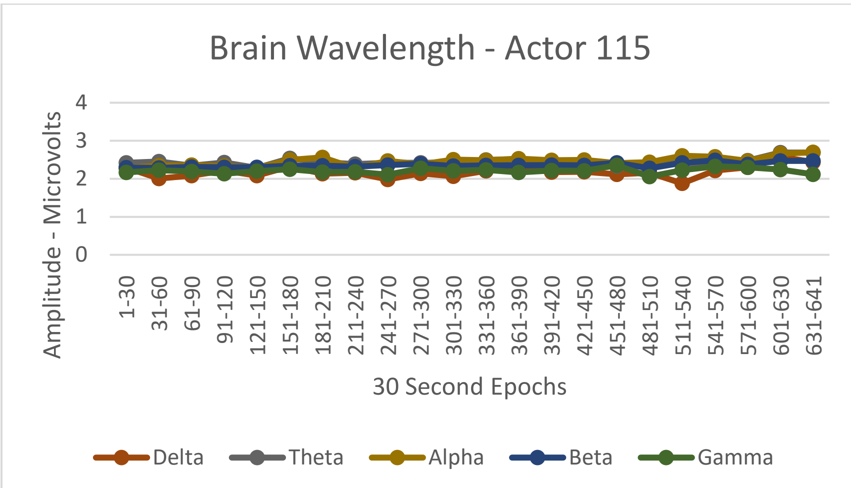

Figure 57 - Actor 115 Brain Wavelength

The actor's workload data was consistent with other sessions. It was very high, approaching 100% at times and is shown in Figure 58. This might have provided additional evidence that his workload increased while controlling the avatar.

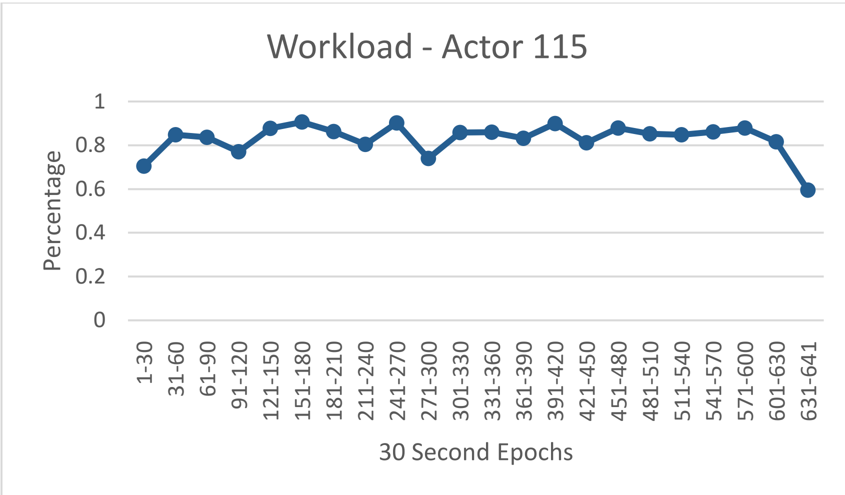

Figure 58 - Actor 115 Workload

Condition 4 – Computer-Controlled Agent

Session 109

Participant 109 was in the computer-controlled agent condition. His/her cognitive state showed a mixed profile as seen in Figure 59. (S)he seemed to move between engagement and disengagement. This participant started the session frustrated because the agent, or AI "misunderstood" his/her dialog and began going into his story before the participant was ready. The AI also repeated information already presented. The participant's frustration seemed to be shown in the influence of distraction.

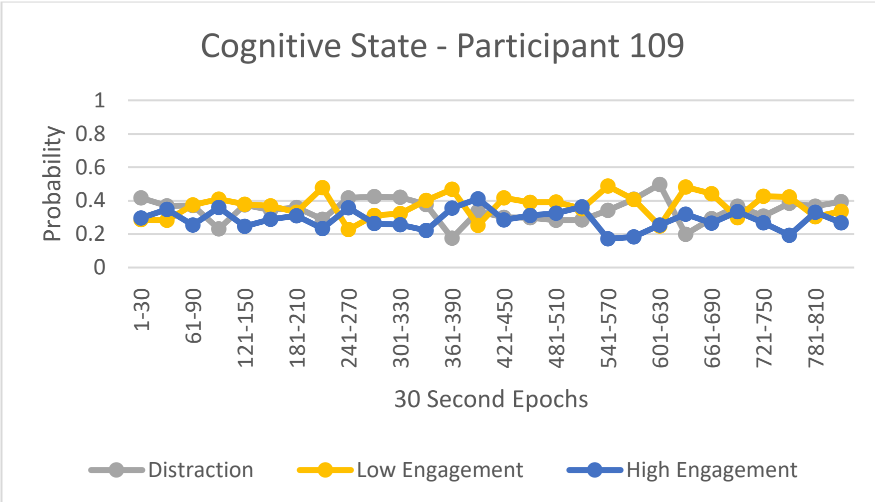

Figure 59 - Participant 109 Cognitive State

The brain wavelength data shown in Figure 60, indicated theta and alpha with a tight coupling. This suggested empathy, emotion and recollection. There were spikes of gamma which were linked to information processing and ideation. This participant experienced sexual assault previously which may have influenced his/her theta waves and indicated recollection.

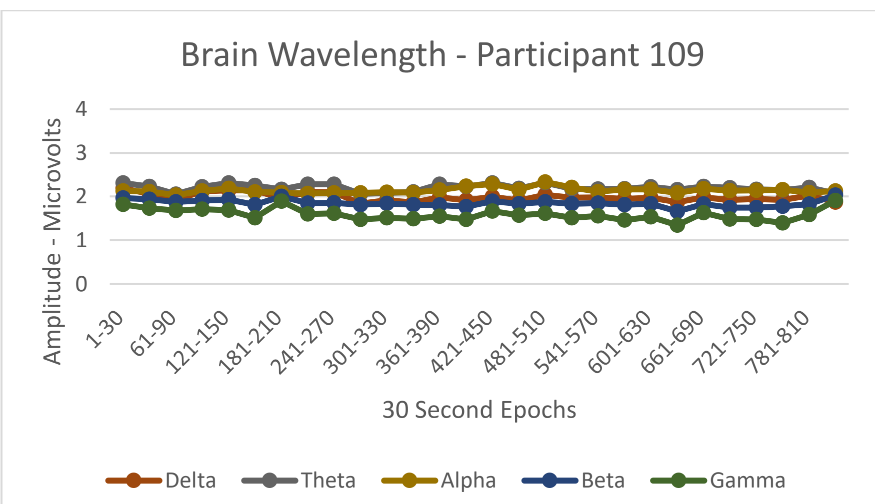

Figure 60 - Participant 109 Brain Wavelength

The workload shown in Figure 61 for Participant 109 was high with very small peaks and valleys that indicated when (s)he was listening rather than speaking. There was no actor data in this condition.

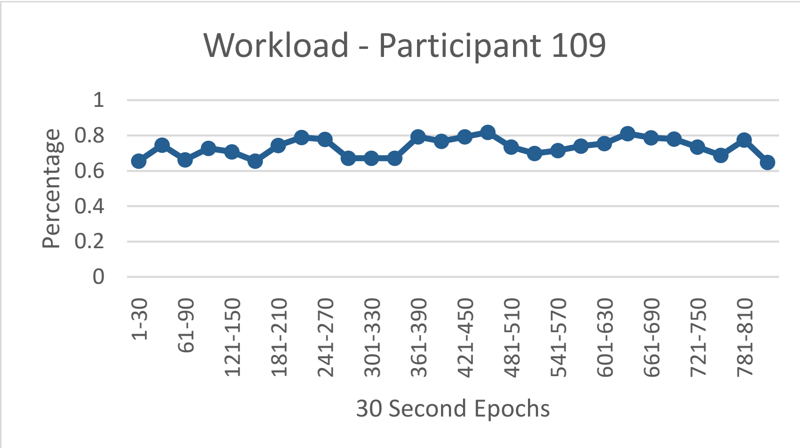

Figure 61 - Participant 109 Workload

Session 116

Participant 116 showed a high probability of high engagement in the beginning of the interchange, but it waned near the end. The video of this participant showed that (s)he was falling asleep during some of the longer agent monologues. This was also clear in the graph in Figure 62 where the drowsy state increased as the session ended and sleep onset rose. These values (drowsy and sleep onset) were removed from the graphs of all other participants because those values had between zero and 0.1 probability in the other graphs.

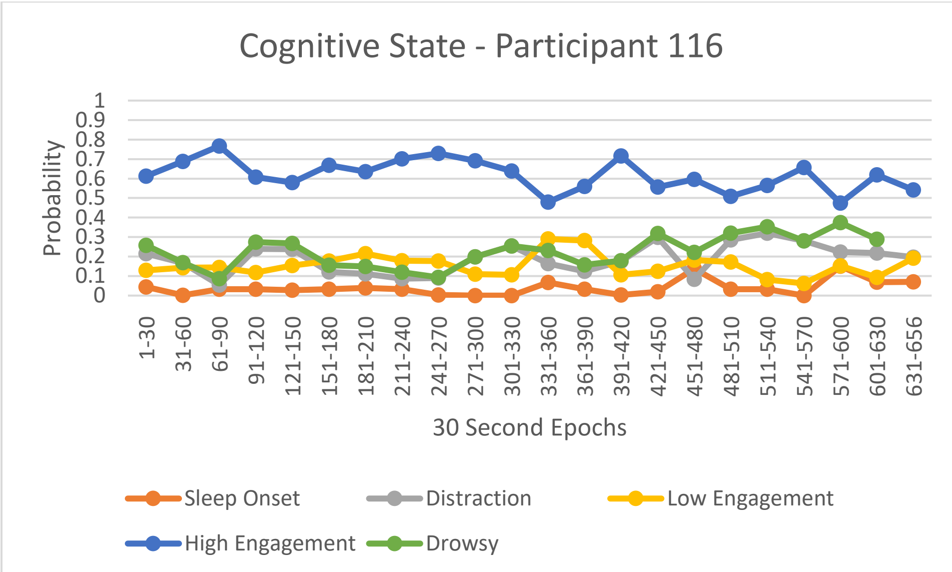

Figure 62 - Participant 116 Cognitive State

Brain wavelength was consistent with other participants and was shown in Figure 63. Alpha and theta had the greatest amplitude followed by gamma. This was associated with emotions and empathy.

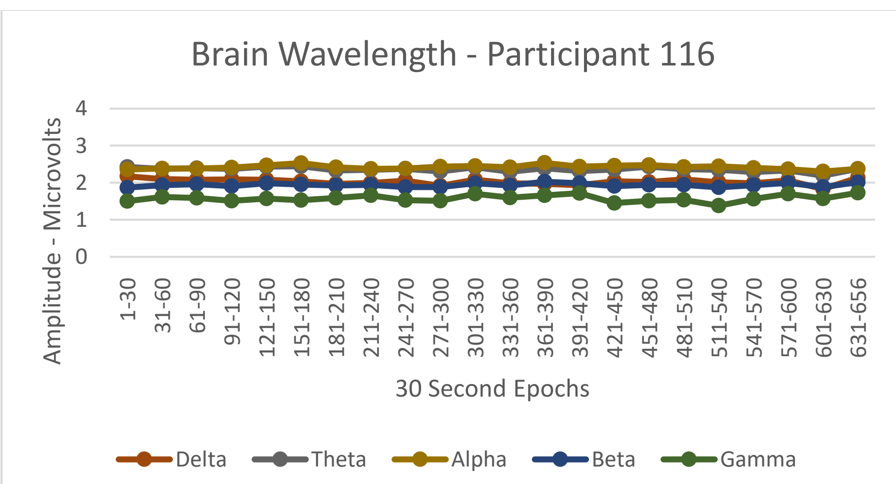

Figure 63 - Participant 116 Brain Wavelength

This participant's workload was fairly neutral until (s)he got closer to the end of the session where the agent's monologues ran long and the participant was struggling to stay awake. This was shown the graph in Figure 64.

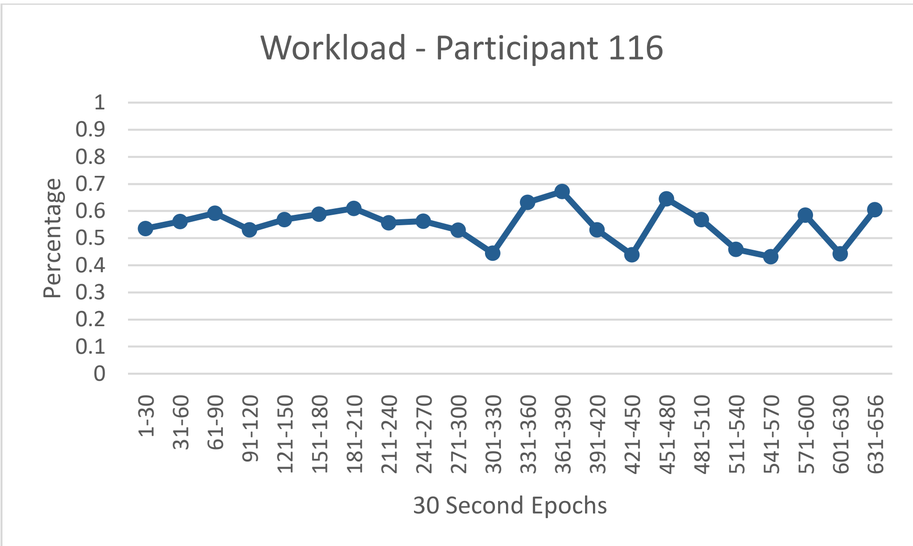

Figure 64 - Participant 116 Workload

Condition 5 – Text

Session 108

In the text condition, the participants had the additional task of translating thought to text then waiting for and interpreting the actor's responses. The results were shown in the cognitive state graph in Figure 65. During this session there was a high probability of low engagement. There were long gaps in response time, since the actor was not a fast typist. The participant stated that (s)he did not know if the gaps were expressions of emotion or slow typing. The participant's engagement ebbed and flowed while distraction seemed to fluctuate with wait times.

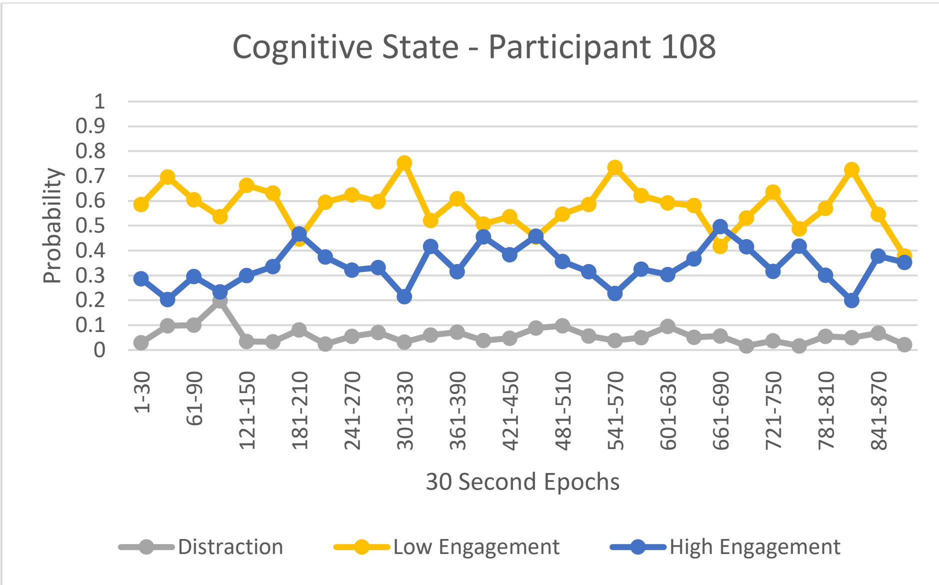

Figure 65 - Participant 108 Cognitive State

The graph shown in Figure 66, showed that the brain wavelengths were tightly coupled with theta and alpha having the highest amplitude. The amplitude of the delta wavelength was also among the highest amplitudes. Delta wavelengths were associated with sleep, suggesting fatigue on the part of the participant.

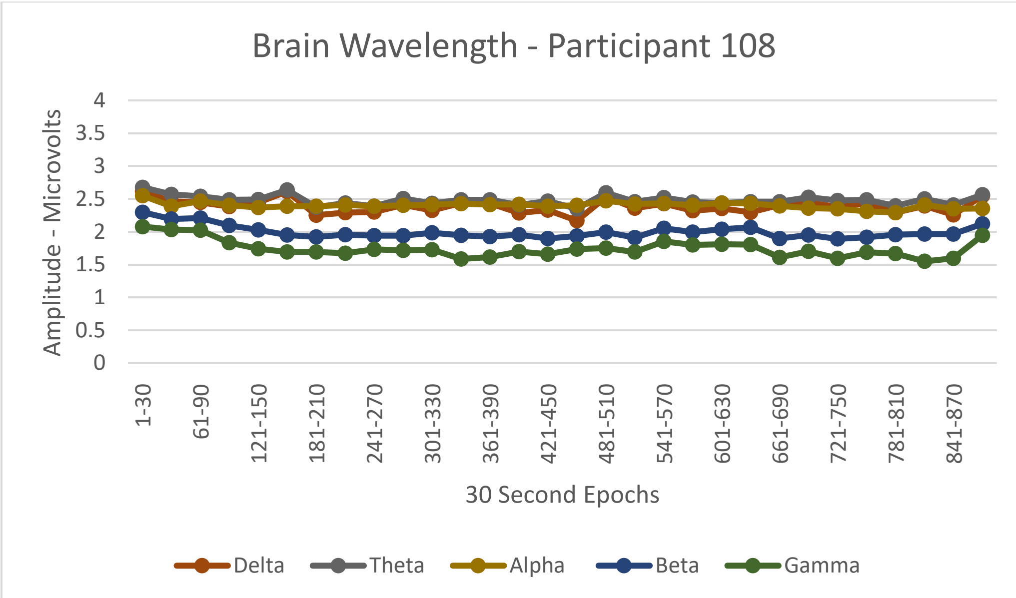

Figure 66 - Participant 108 Brain Wavelength

Workload for participant 108, started out high then dropped during the dialog as can be seen in Figure 67. This occurred as wait-times increased for the responses from the actor.

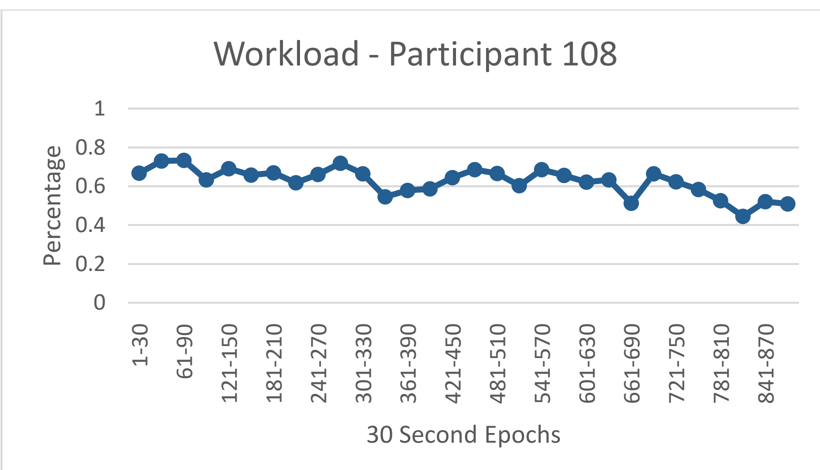

Figure 67 - Participant 108 Workload

The actor's EEG data for the same session follows. The cognitive state of the actor was inconsistent with previous graph readings in this area as was shown in Figure 68. This appeared to depict the frustration the actor was experiencing as he attempted to tell his story while translating his message via text. Since the actor did not type quickly, the eight to ten-minute interval was not enough time to relay the story to the participant. The graph showed high levels of distraction which appeared to be directly related to the taxing activity of typing.

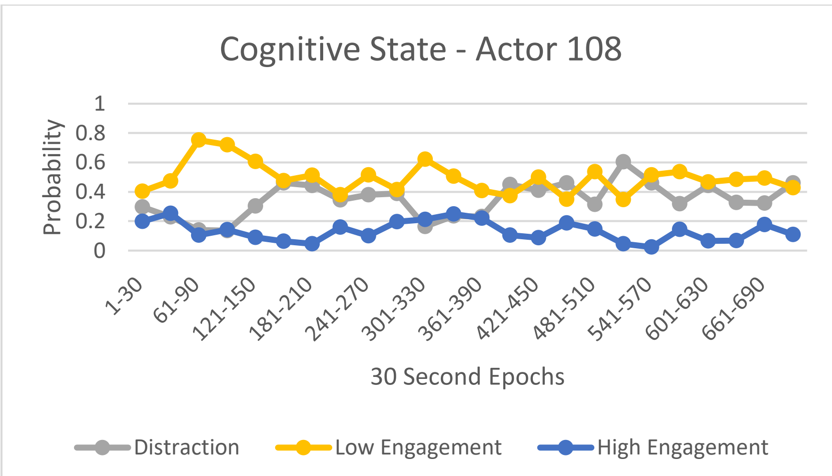

Figure 68 - Actor 108 Cognitive State

While there was a gap in the brain wavelength data shown in Figure 69, it was clear that the brain wavelengths showed a consistent pattern with previous actor profiles. Theta and alpha amplitudes were high and we saw some level of beta expressed in higher amplitudes, which might have been related to caffeine consumption.

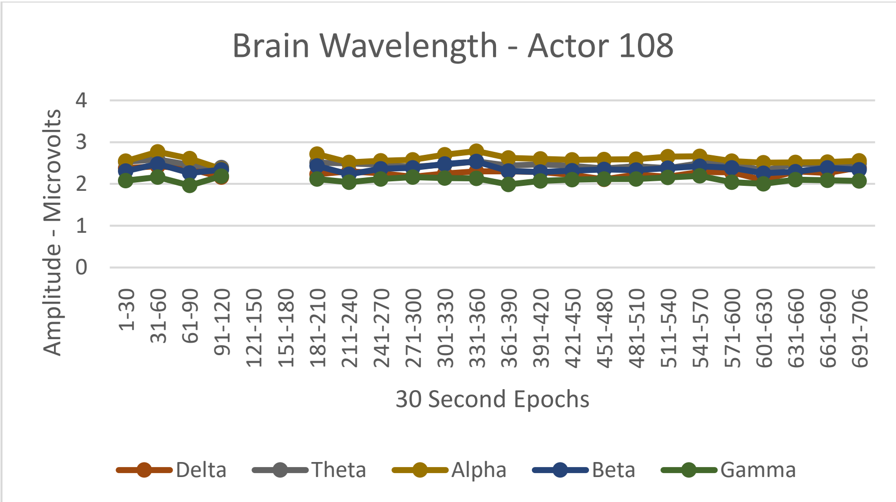

Figure 69 - Actor 108 Brain Wavelength

The same gap in data was apparent in workload shown in Figure 70. Workload was even higher than normal for the actor, which was likely associated with the additional task of typing his story.

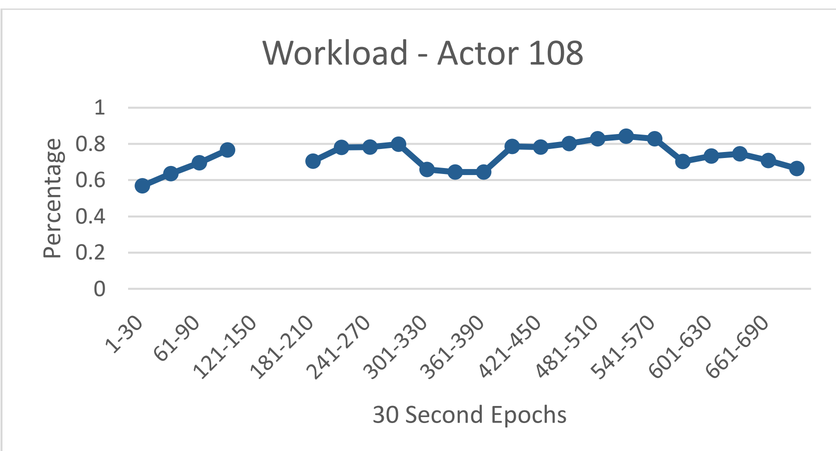

Figure 70 - Actor 108 Workload

Session 114

The cognitive state graph for this session is shown in Figure 71. This participant showed likelihood of high engagement with periods of distraction that might be caused by the time to translate thought to text, wait for the actor's response, and decode it. This graph showed a similar pattern to the previous pattern of cognitive state for the actor. This might have been because the actor had help in typing this time.

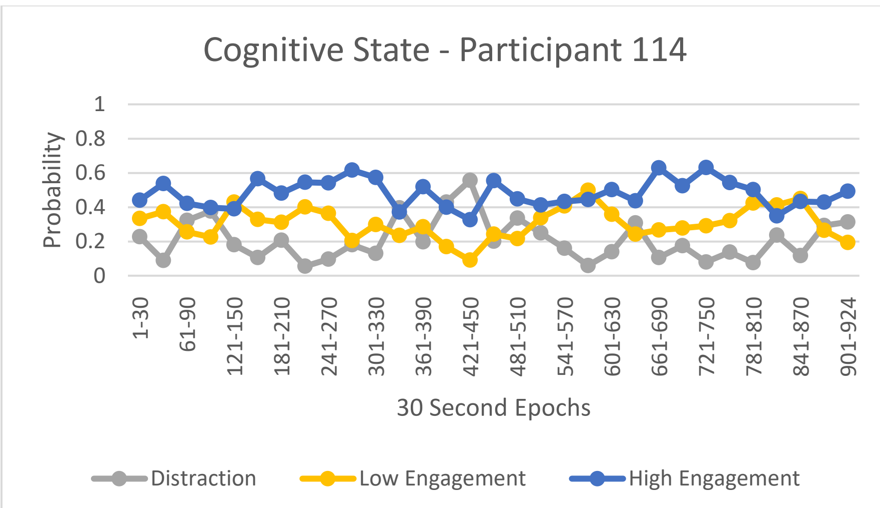

Figure 71 - Participant 114 Cognitive State

The brain wavelength graph, in Figure 72, showed a recognizable pattern, consistent with most participants with alpha and theta having higher amplitudes. Delta was also seen, similar to the previous text condition, suggesting some fatigue in processing the text.

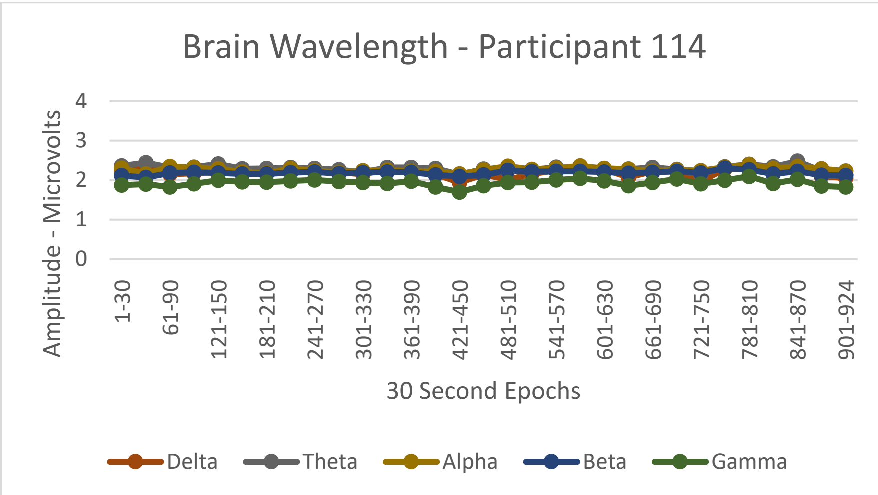

Figure 72 - Participant 114 Brain Wavelength

The participant's workload in Figure 73, had many peaks and valleys, but was not high. There was only one area where the peak was well above 5.5 near the end of the interchange.

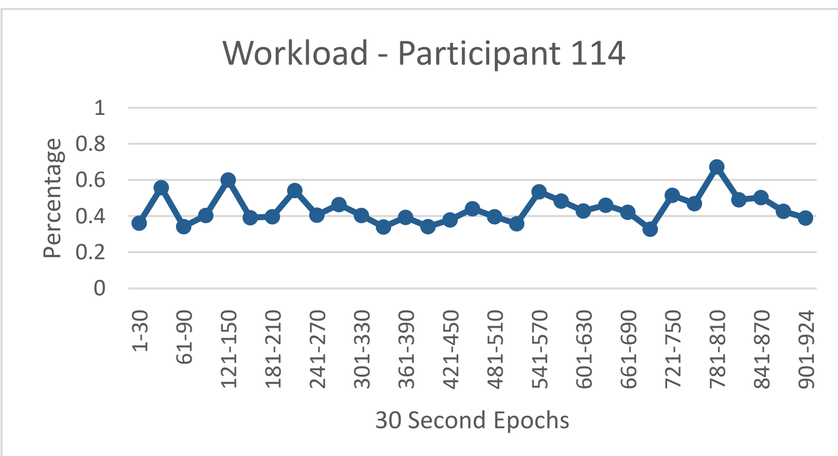

Figure 73 - Participant 114 Workload

The actor's EEG data for the same session follows. The cognitive state of the actor was not consistent with previous sessions as can be seen in Figure 74. Data anomalies appeared to exist in this graph. This chart suggested that the actor was sleeping, but since he was working with someone who was helping with typing, it was clear he was fully engaged as the two discussed responses.

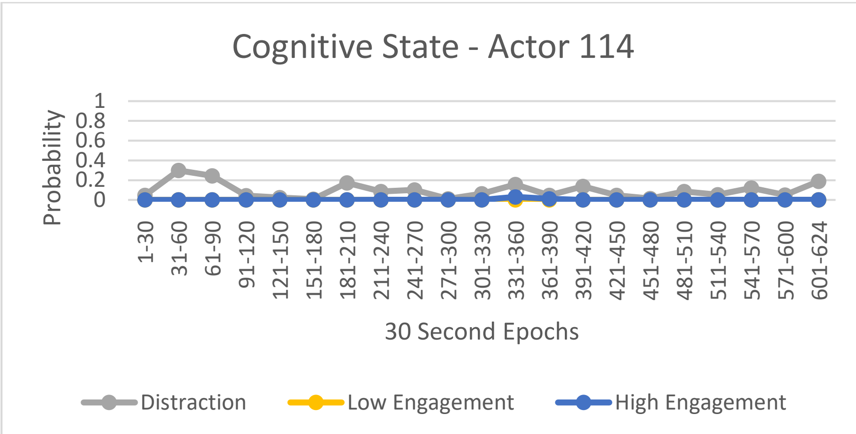

Figure 74 - Actor 114 Cognitive State

The actor's cognitive state was consistent with theta and alpha being primary, but delta had a much more significant role in this graph than in previous cases. This could be seen in the data in Figure 75.

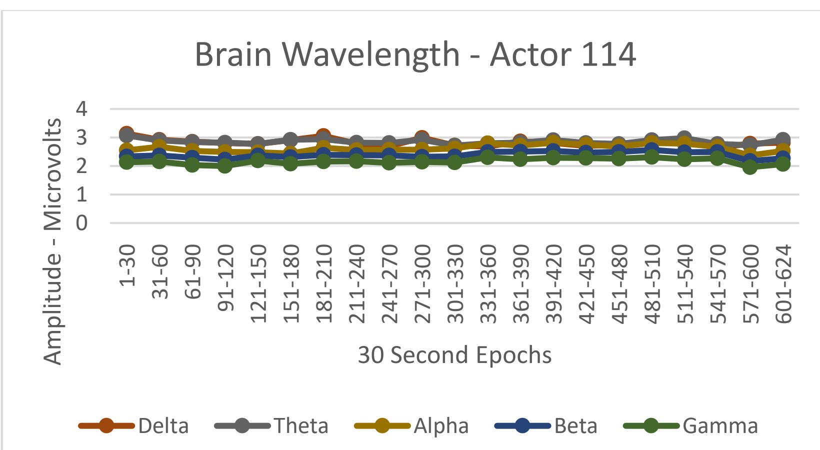

Figure 75 - Actor 114 Brain Wavelength

The actor's workload data shown in Figure 76, showed high workload, even higher than the actor's normal workload. It was possible that the translation from text might have driven workload higher, but the fact that the actor had typing help in this case, suggest otherwise.

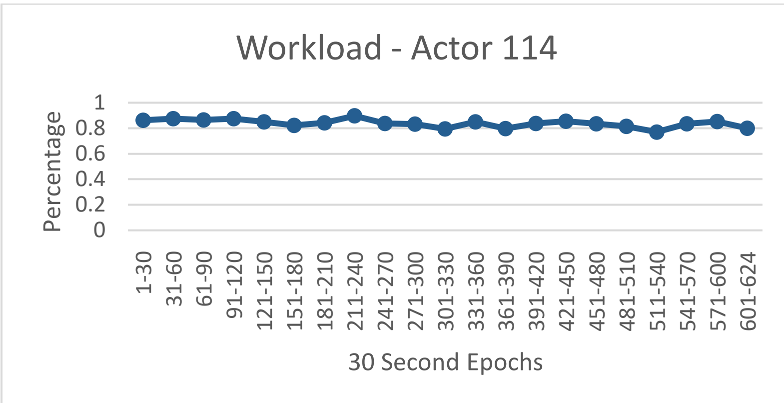

Figure 76 - Actor 114 Workload

Heart Rate

The B-Alert Systems also collected heart rate data on both the participants and the actor. Heart rate was collected to determine if it had any value in assessing if a participant had the sense that the dialog partner had agency. Data on heart rate were shown below by condition. Research (Kazmi, et al., 2016) shows that heart rate and heart rate variability have an inverse relationship. It was unclear what this data offered in answering the research questions but was shown here for information only. Further discussion on heart rate data is in CHAPTER FIVE.

Condition 1 - Face-to-Face

Session 106

In the face-to-face condition, the first participant's heart rate graph is shown in Figure 77.

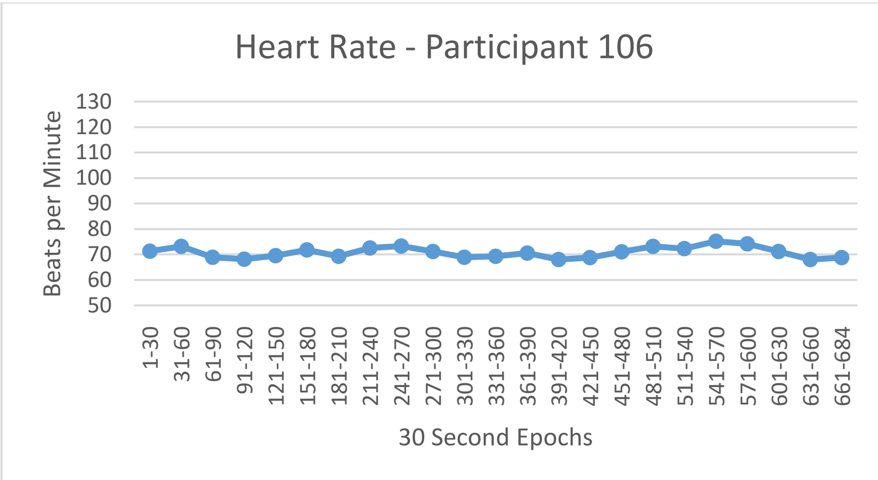

Figure 77 - Participant 106 Heart Rate

The same heart rate data as the participant above, was collected on the actor. The actor's heart rate data for the same session follows. The heart rate data from the actor is shown in Figure 78.

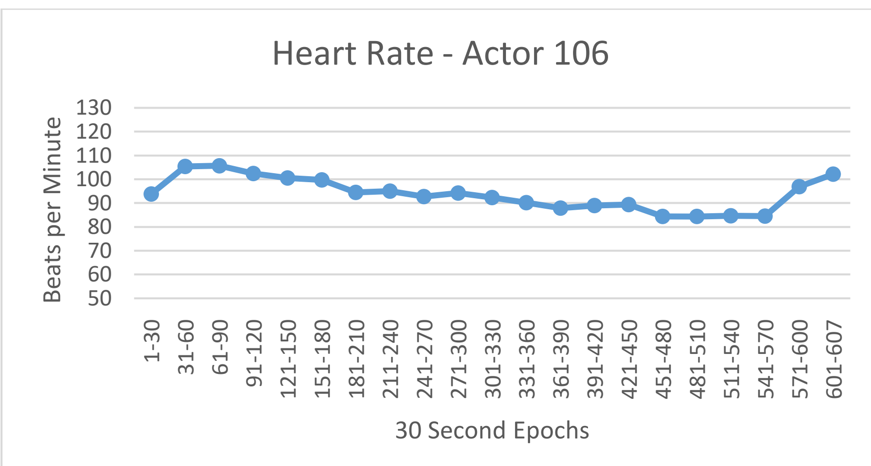

Figure 78 - Actor 106 Heart Rate

Session 110

The heart rate data for participant 110 is shown in the graph in Figure 79.

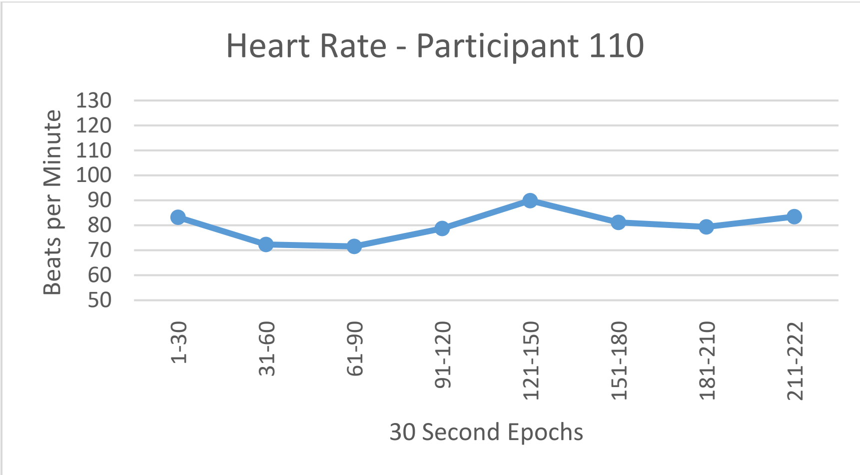

Figure 79 - Participant 110 Heart Rate

The actor's heart rate data for the same session follows. The heart rate of the actor is shown in Figure 80.

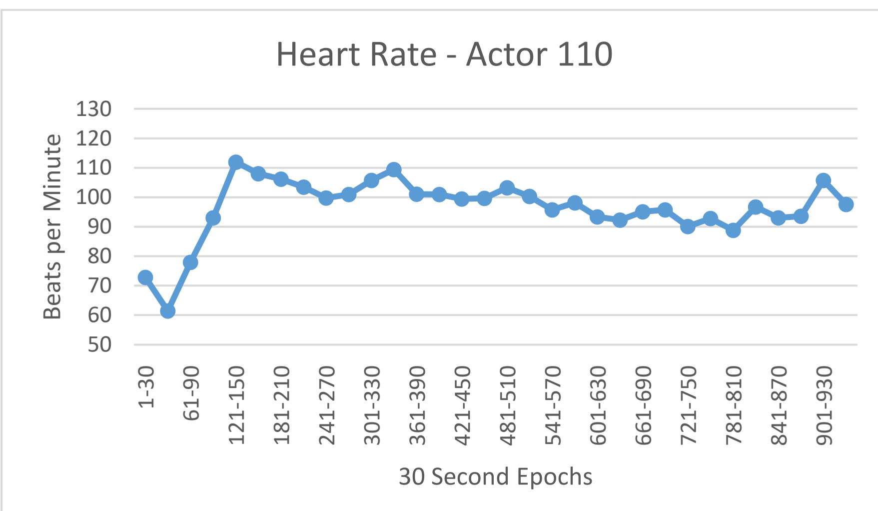

Figure 80 - Actor 110 Heart Rate

Session 111

The heart rate data for participant 111 is shown in the graph in Figure 81.

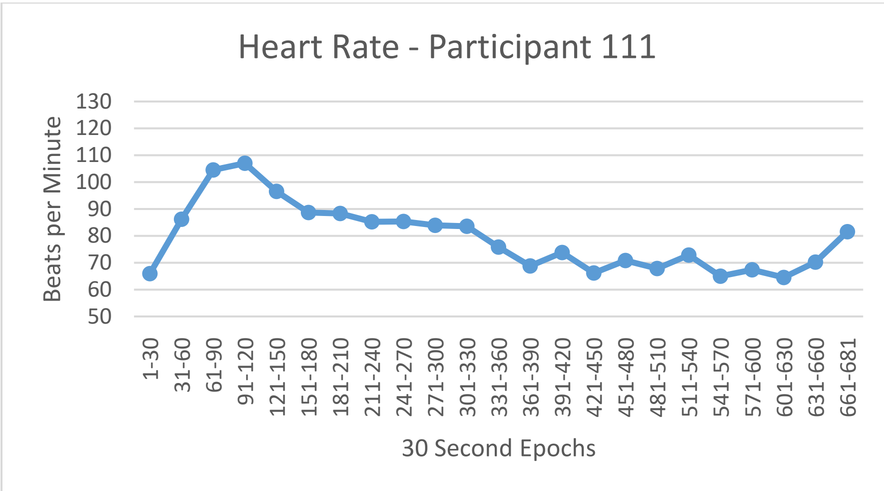

Figure 81 - Participant 111 Heart Rate

The actor's heart rate data for the same session follows. The heart rate data of the actor is shown in Figure 82.

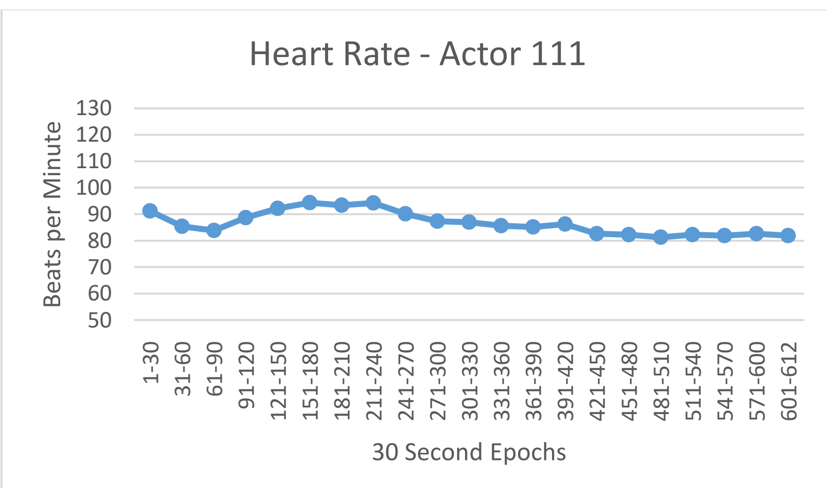

Figure 82 - Actor 111 Heart Rate

Condition 2 – Video

Session 105

In the video condition, the first session heart rate data is shown in Figure 83.

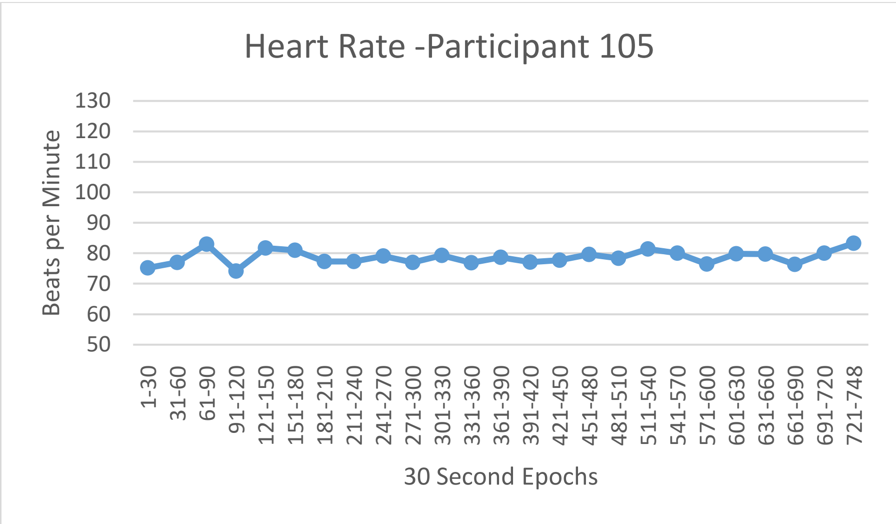

Figure 83 - Participant 105 Heart Rate

The actor's data for the same session follows. The heart rate of the actor is shown in Figure 84.

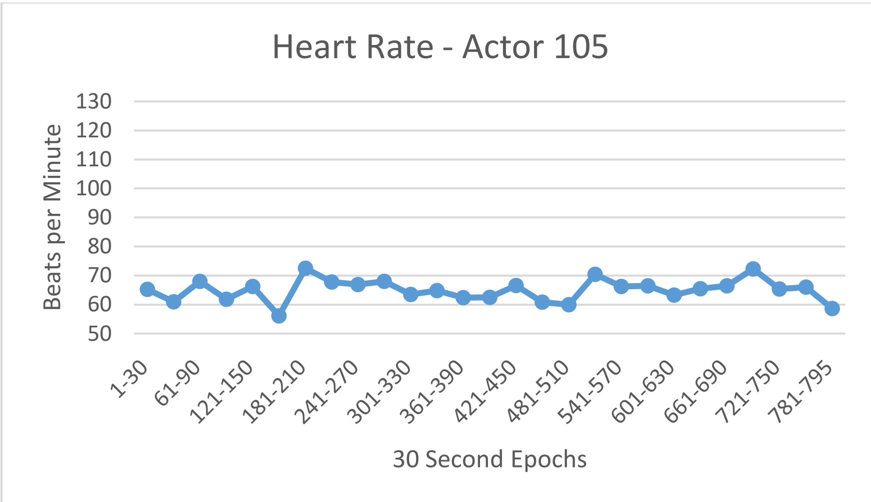

Figure 84 - Actor 105 Heart Rate

Session 113

The heart rate graph for participant 113 is shown in Figure 85.

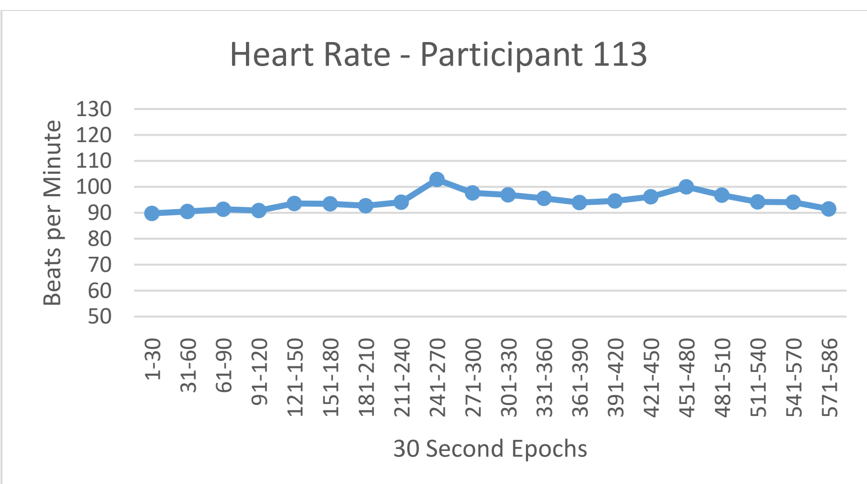

Figure 85 - Participant 113 Heart Rate

The actor's data for the same session follows. The heart rate of the actor is shown in Figure 86.

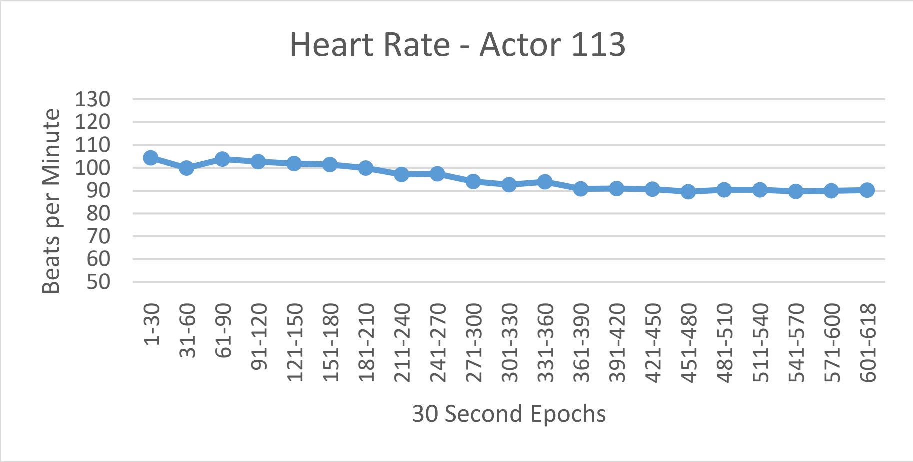

Figure 86 - Actor 113 Heart Rate

Condition 3 – Human-Controlled Avatar

Session 107

In the human-controlled avatar condition, the heart rate graph for participant 107 is shown in Figure 87.

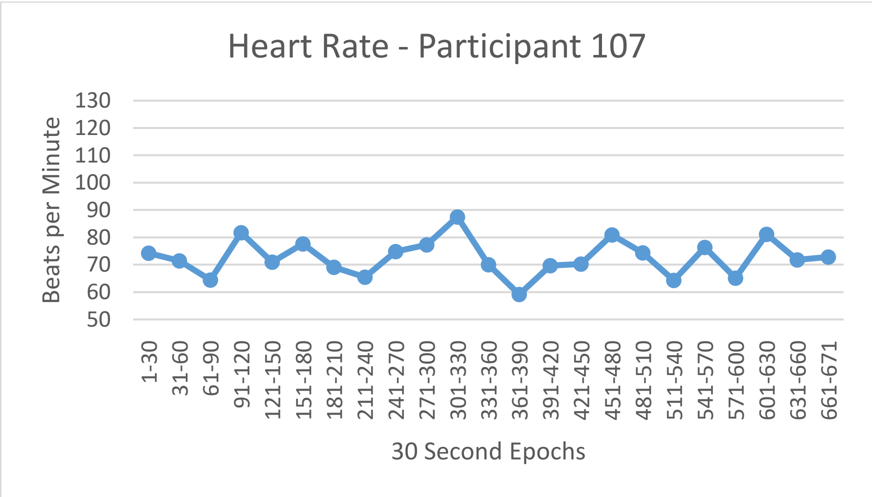

Figure 87 - Participant 107 Heart Rate

The actor's data for the same session follows. The heart rate of the actor is shown in Figure 88.

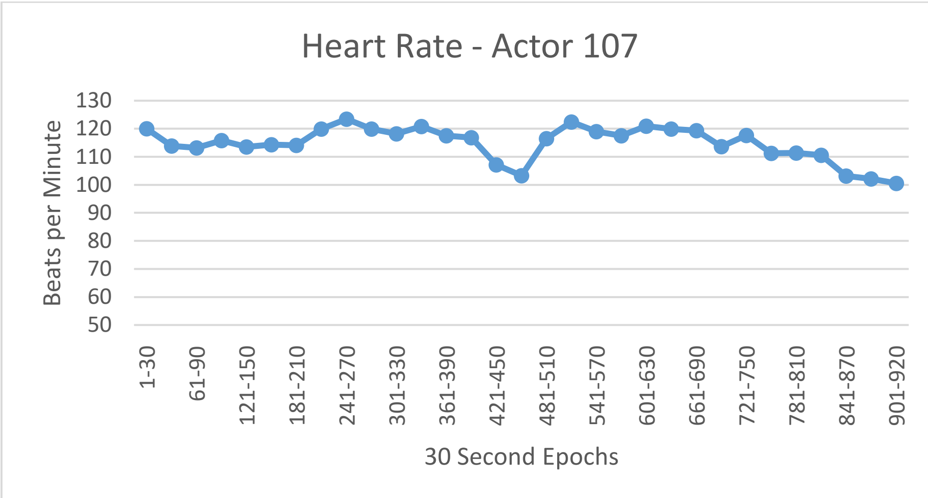

Figure 88 - Actor 107 Heart Rate

Session 112

The heart rate data for participant 112 is shown in Figure 89.

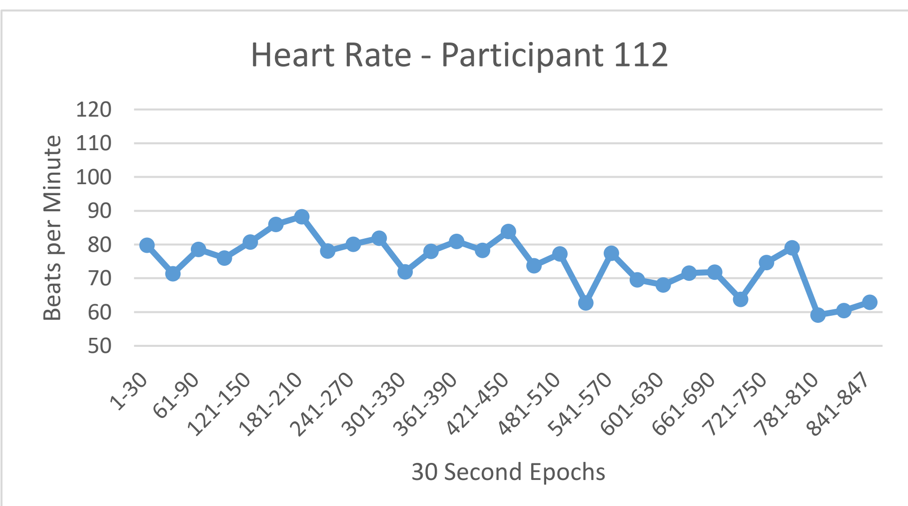

Figure 89 - Participant 112 Heart Rate

The actor's data for the same session follows in Figure 90.

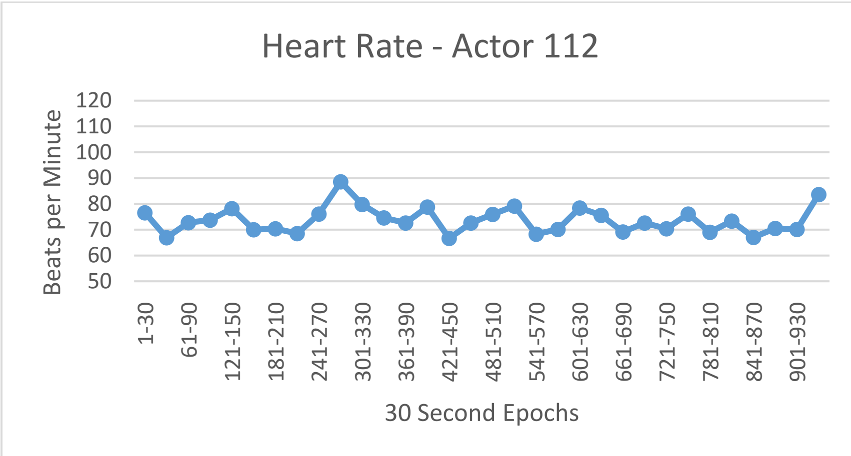

Figure 90 - Actor 112 Heart Rate

Session 115

The heart rate graph for participant 115 is shown in Figure 91.

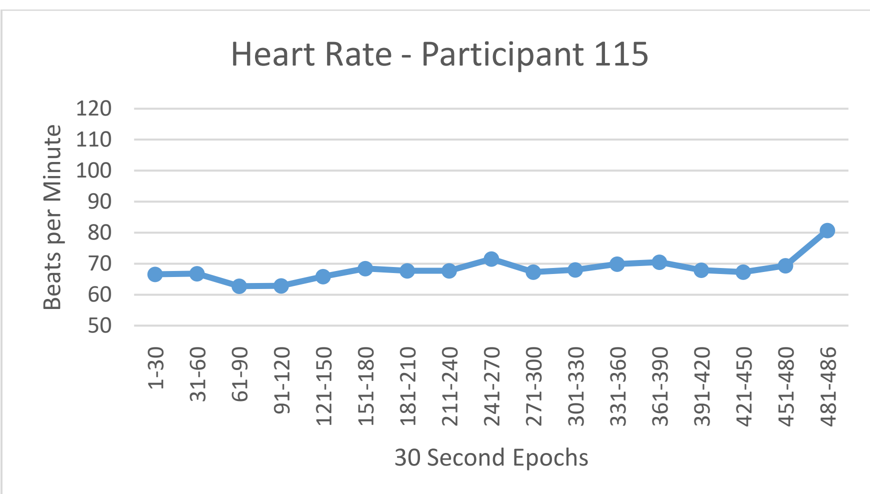

Figure 91 - Participant 115 Heart Rate

The actor's data for the same session follows in Figure 92.

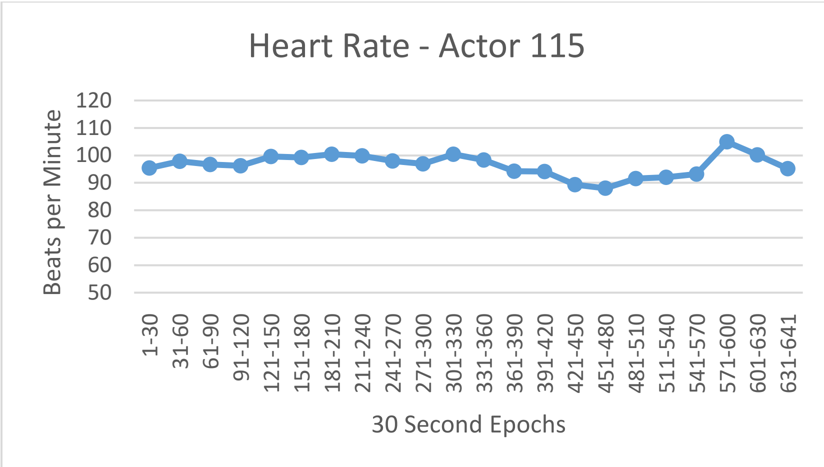

Figure 92 - Actor 115 Heart Rate

Condition 4 – Computer-Controlled Agent

Session 109

In the computer-controlled agent condition, the first participant's heart rate graph is shown in Figure 93.

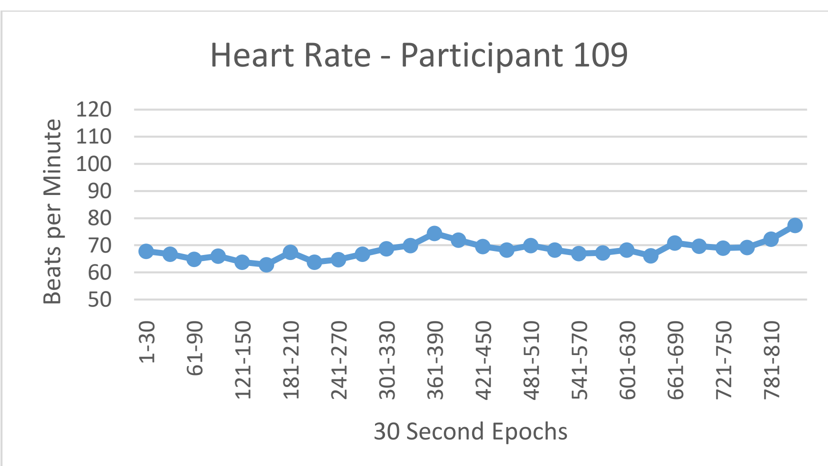

Figure 93 - Participant 109 Heart Rate

Session 116

The heart rate graph for participant 116 is shown in Figure 94.

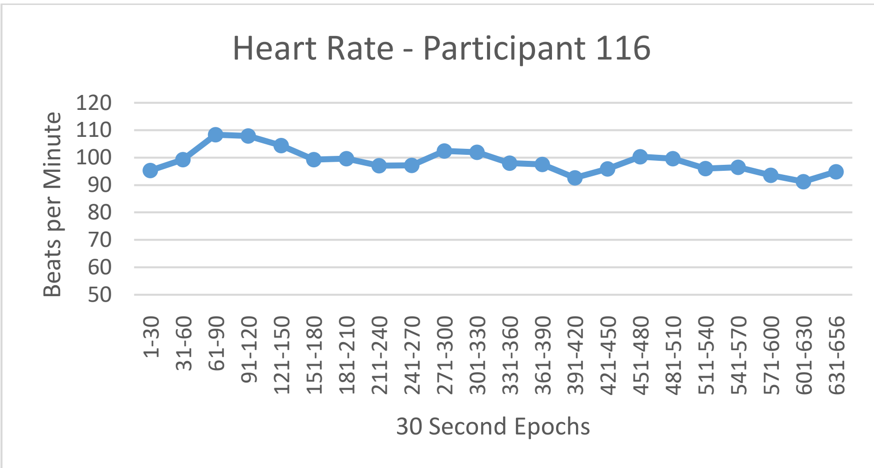

Figure 94 - Participant 116 Heart Rate

Condition 5 – Text

Session 108

In the text condition, the heart rate graph for participant 108 is shown in Figure 95.

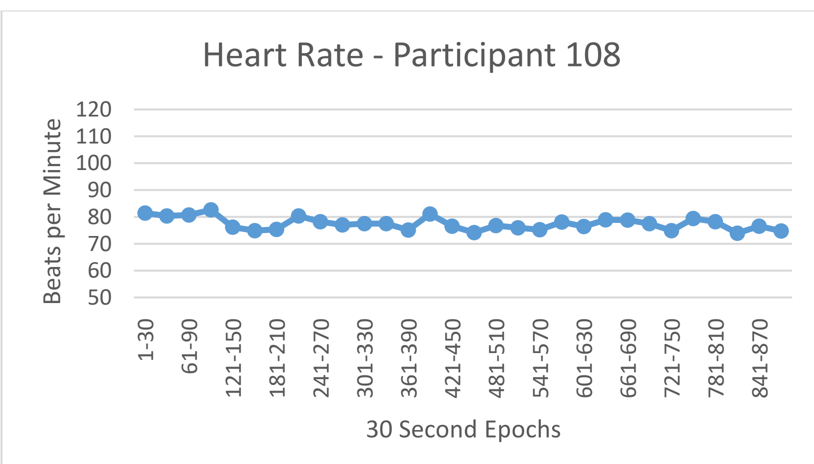

Figure 95 - Participant 108 Heart Rate

The actor's data for the same session is shown in Figure 96.

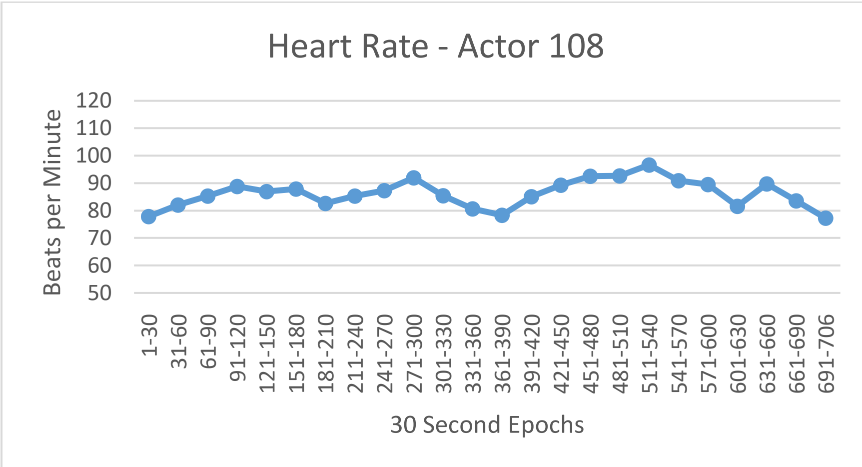

Figure 96 - Actor 108 Heart Rate

Session 114

The heart rate graph for participant 114 is shown in Figure 97.

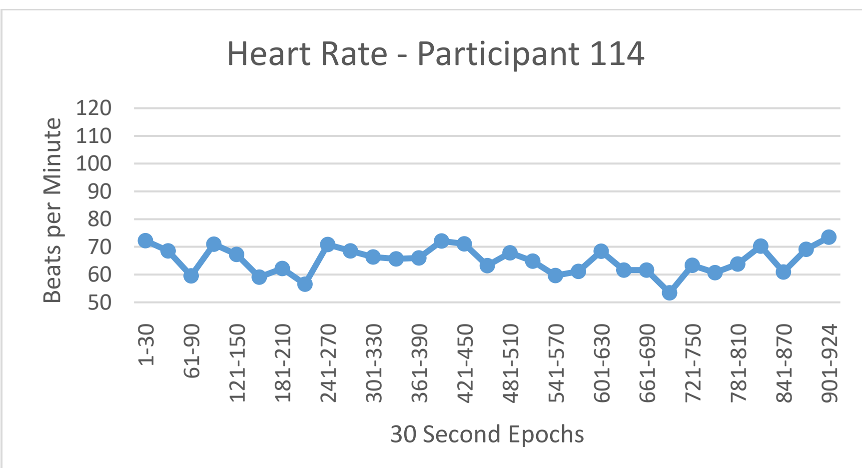

Figure 97 - Participant 114 Heart Rate

The actor's data for the same session is shown in Figure 98.

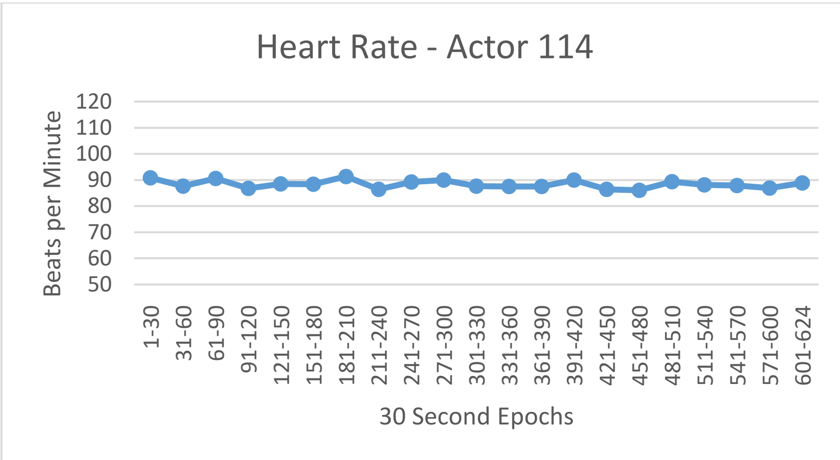

Figure 98 - Actor 114 Heart Rate

Heart Rate Variability or Inter-Beat Interval

Heart rate variability, otherwise known as Inter-Beat Interval (IBI) was another physiological measure that was being assessed to determine if it was a reliable indicator of an interactive partner seeming to have agency. It was unclear that the IBI values provided any insight into the sense of agency. Despite that, the data were presented below in order of condition. It will be discussed further in CHAPTER FIVE.

Condition 1 - Face-to-Face

Session 106

In the face-to-face condition, the first participant's data are shown in the IBI graph in Figure 99.

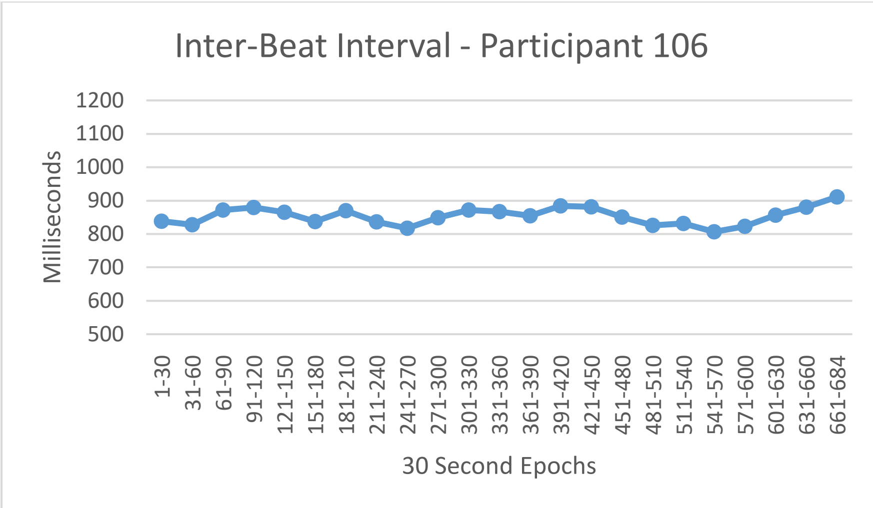

Figure 99 - Participant 106 Inter-Beat Interval

The same IBI data as the participant above, was collected on the actor. The actor's IBI data for the same session follows in Figure 100.

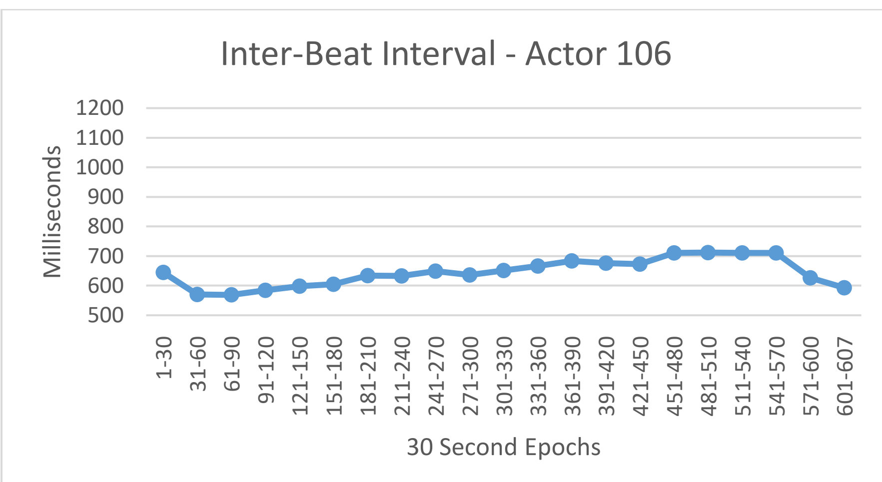

Figure 100 - Actor 106 Inter-Beat Interval

Session 110

IBI information for participant 110 is shown in Figure 101.

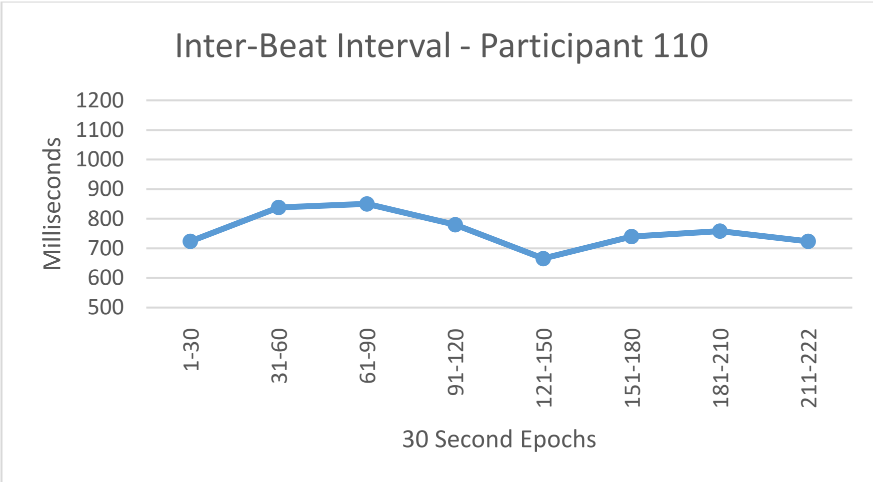

Figure 101 - Participant 110 Inter-Beat Interval

The actor's IBI data for the same session is shown in Figure 102.

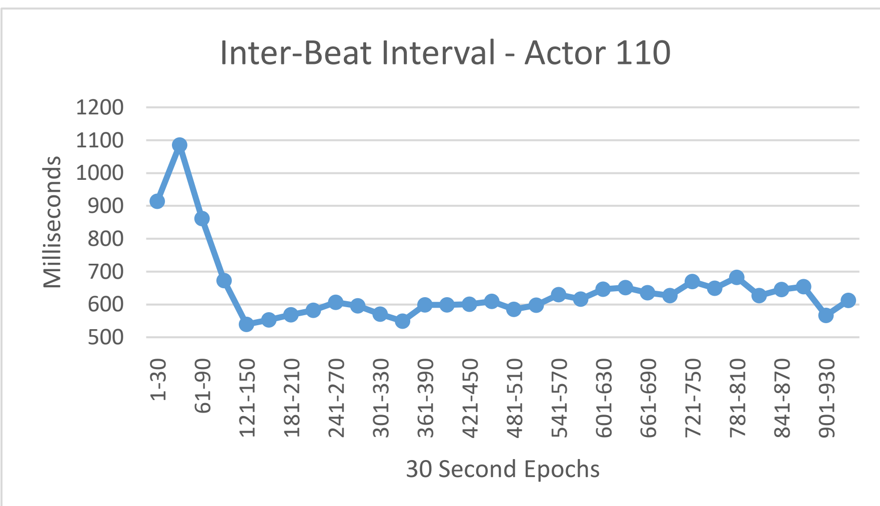

Figure 102 - Actor 110 Inter-Beat Interval

Session 111

The IBI for participant 111 is shown in Figure 103.

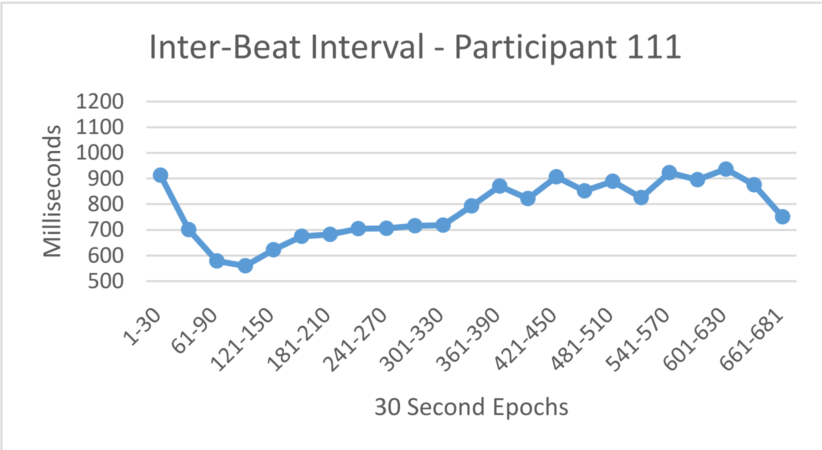

Figure 103 - Participant 111 Inter-Beat Interval

The actor's IBI data for the same session is shown in Figure 104.

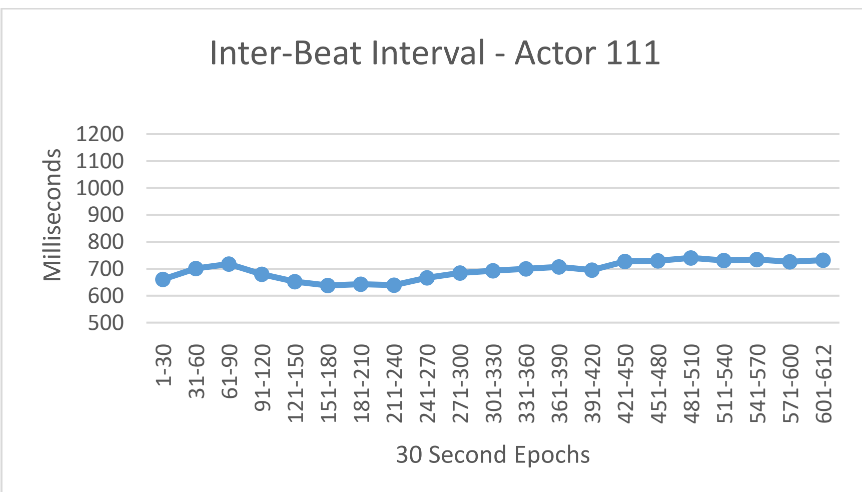

Figure 104 - Actor 111 Inter-Beat Interval

Condition 2 – Video

Session 105

In the video condition, the IBI for participant 105 is shown in Figure 105.

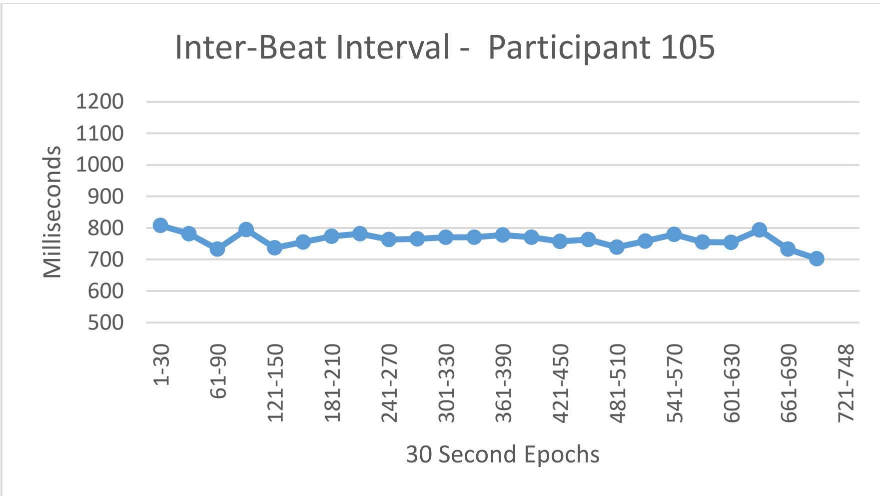

Figure 105 - Participant 105 Inter-Beat Interval

The actor's data for the same session is shown in Figure 106

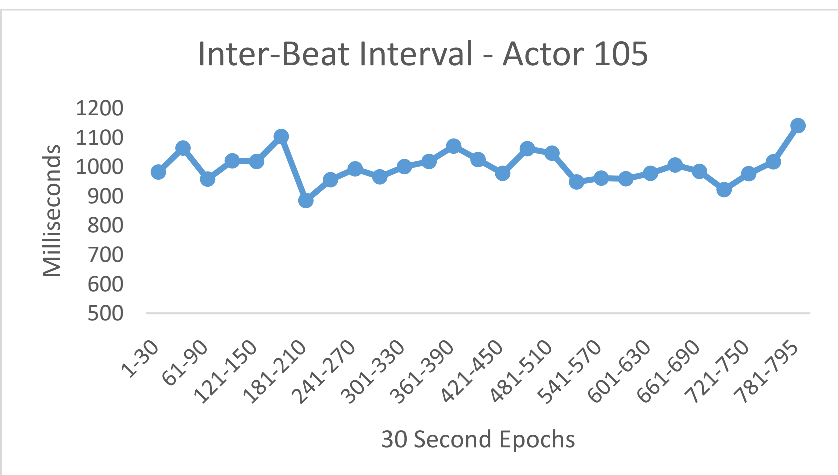

Figure 106 - Actor 105 Inter-Beat Interval

Session 113

The IBI graph for participant 113 is shown in Figure 107.

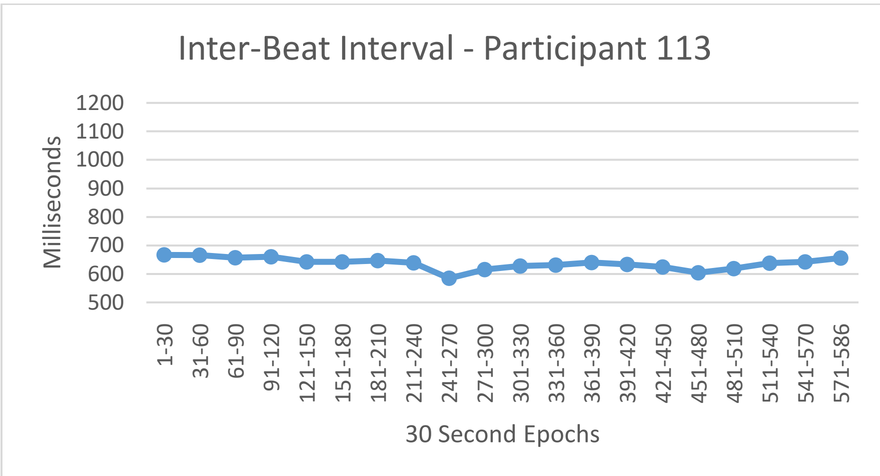

Figure 107 - Participant 113 Inter-Beat Interval

The actor's data for the same session is shown in Figure 108.

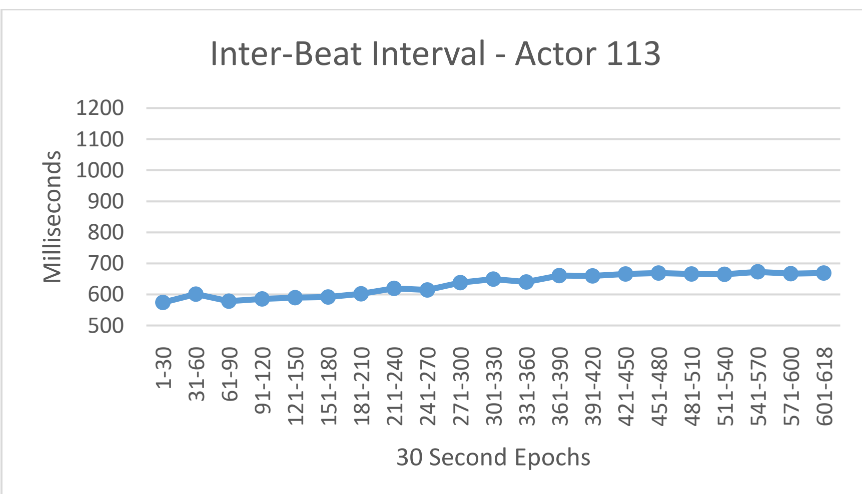

Figure 108 - Actor 113 Inter-Beat Interval

Condition 3 – Human-Controlled Avatar

Session 107

In the human-controlled avatar condition, the IBI graph for participant 107 is shown in Figure 109.

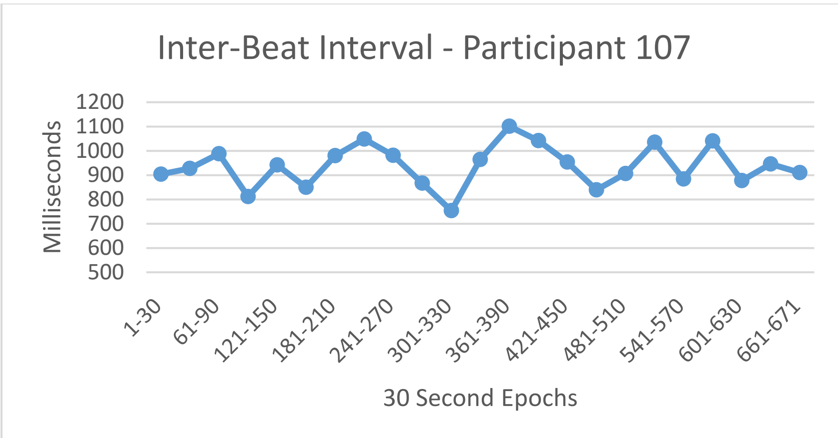

Figure 109 - Participant 107 Inter-Beat Interval

The actor's data for the same session follows. The IBI of the actor is shown in Figure 110.

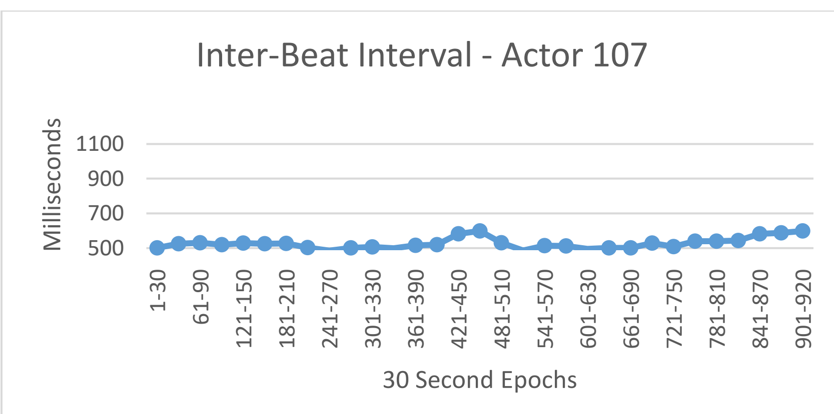

Figure 110 - Actor 107 Inter-Beat Interval

Session 112

The IBI graph for participant 112 is shown in Figure 111.

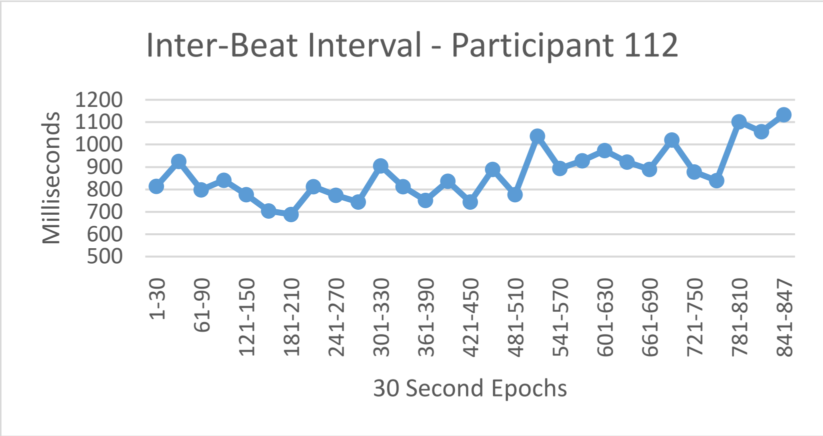

Figure 111 - Participant 112 Inter-Beat Interval

The actor's data for the same session follows. The IBI data of the actor is shown in Figure 112.

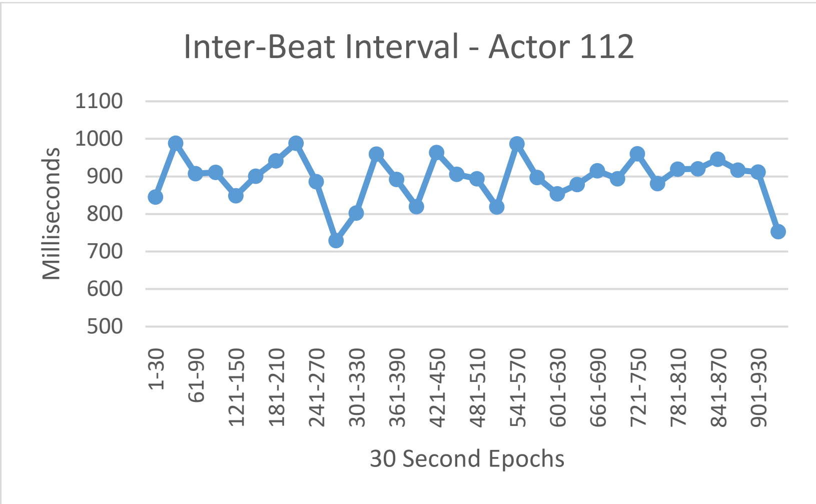

Figure 112 - Actor 112 Inter-Beat Interval

Session 115

The IBI data for participant 115 is shown in Figure 113.

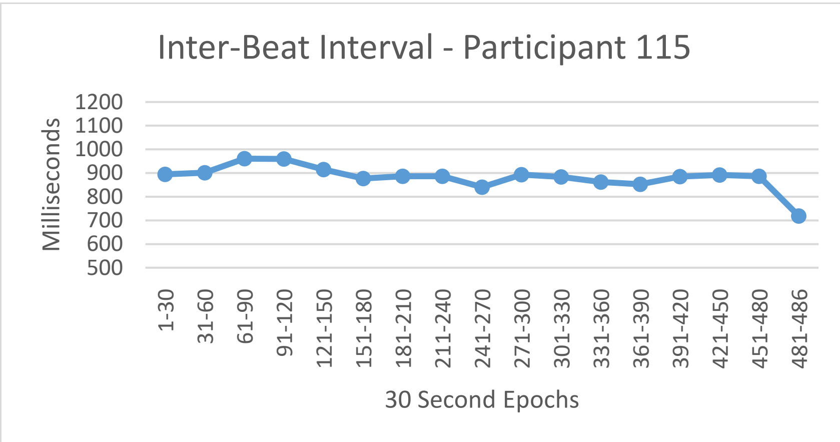

Figure 113 - Participant 115 Inter-Beat Interval

The actor's data for the same session follows. The IBI of the actor is shown in Figure 114.

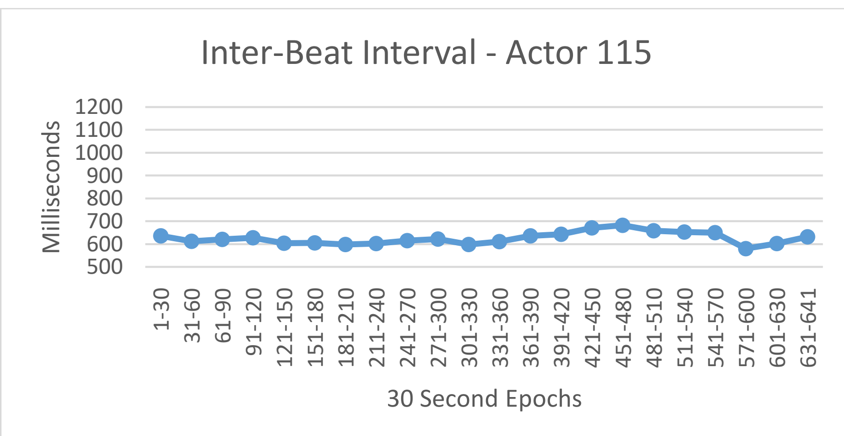

Figure 114 - Actor 115 Inter-Beat Interval

Condition 4 – Computer-Controlled Agent

Session 109

In the computer-controlled agent condition, the IBI graph for participant 109 is shown in Figure 115.

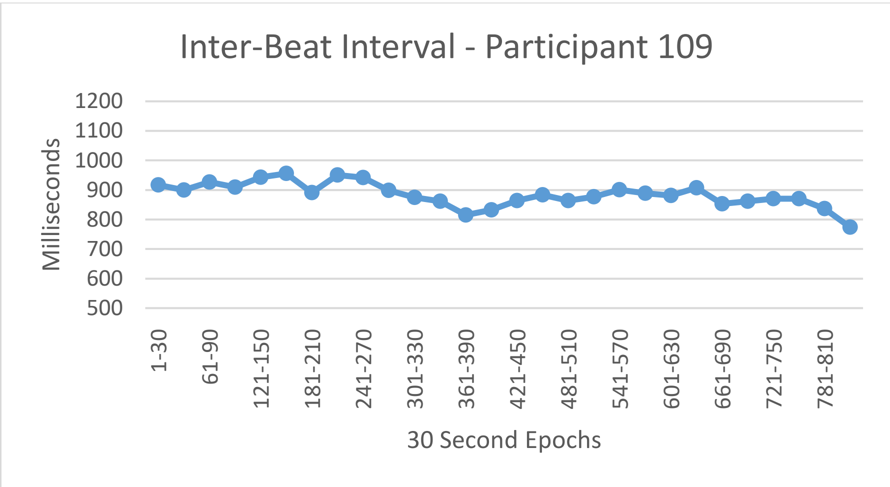

Figure 115 - Participant 109 Inter-Beat Interval

Session 116

The IBI graph for participant 116 is shown in Figure 116.

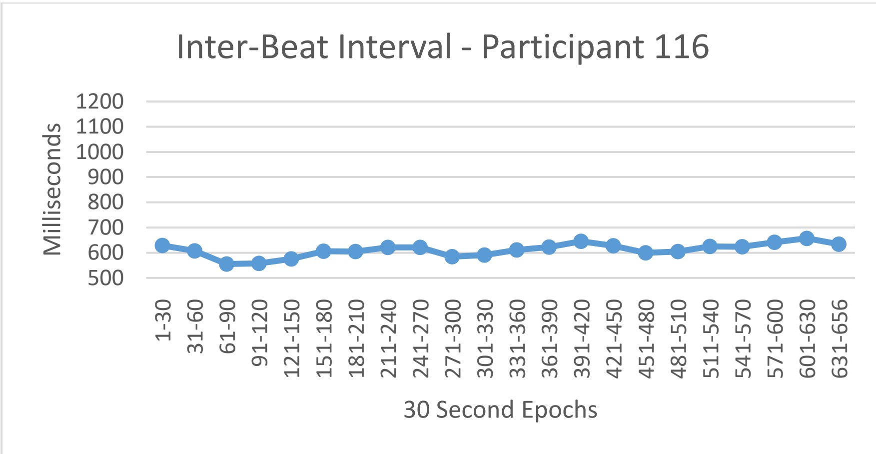

Figure 116 - Participant 116 Inter-Beat Interval

Condition 5 – Text

Session 108

In the text condition, the IBI for participant 108 is shown in Figure 117.

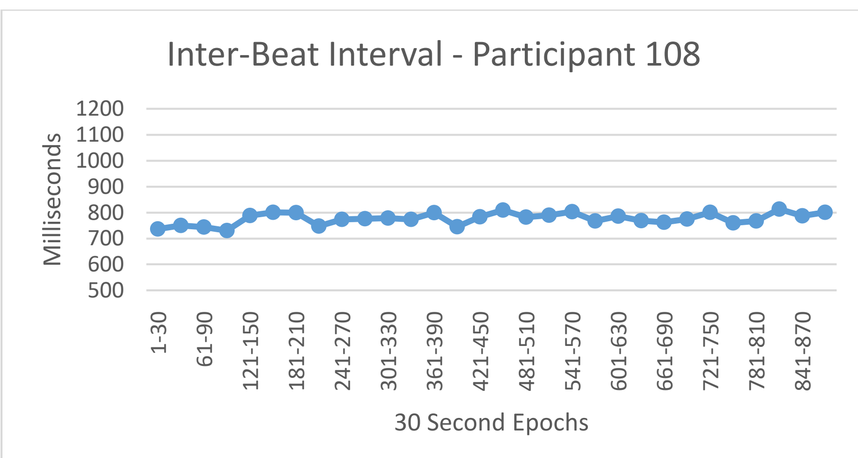

Figure 117 - Participant 108 Inter-Beat Interval

The actor's data for the same session follows. The IBI of the actor is shown in Figure 118.

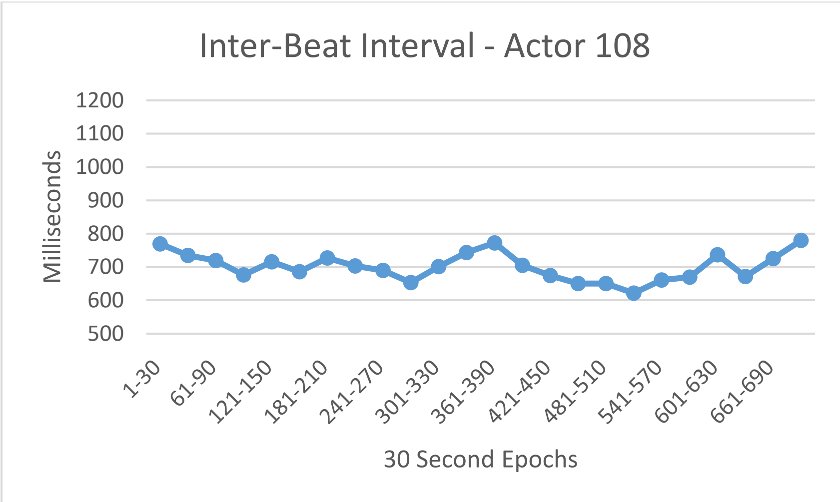

Figure 118 - Actor 108 Inter-Beat Interval

Session 114

The IBI graph for participant 114 is shown in Figure 119.

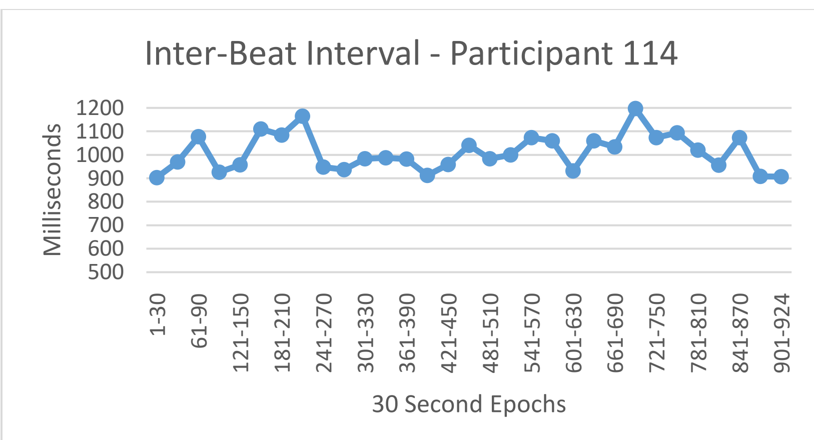

Figure 119 - Participant 114 Inter-Beat Interval

The actor's data for the same session follows. The IBI of the actor is shown in Figure 120.

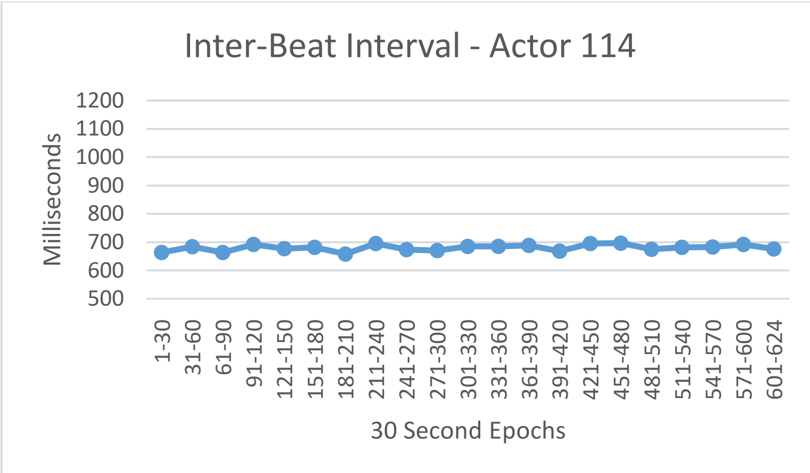

Figure 120 - Actor 114 Inter-Beat Interval

Electrodermal Activity

EDA was measured using the Affectiva Q GSR Sensor. The EDA data were challenging to assess. The first challenge was to synchronize the start time with the video, EEG and dialog events. Values were also wildly inconsistent from one participant to another, which could be accounted for by individual differences, or could be issues with the equipment. The Q-Sensor was no longer supported by Affectiva, so there was no technical support to determine if the sensors were actually collecting data appropriately.

The intent of using EDA was to determine if sufficient data were picked up by the sensors to indicate that the actor had a sense of their dialog partner having agency. Unfortunately, quality data were very limited for these sessions with only three results appearing to provide reasonable data. That data was provided below with a discussion on what was seen.

Condition 1 - Face-to-Face

Session 110

Session 110 was the only sessions with meaningful EDA data in this condition. The data are presented below.

The EDA data for participant 110 was shown in Figure 121. The initial spike indicated the start of the session. The following spike appeared to be noise and could not be associated with any particular event.

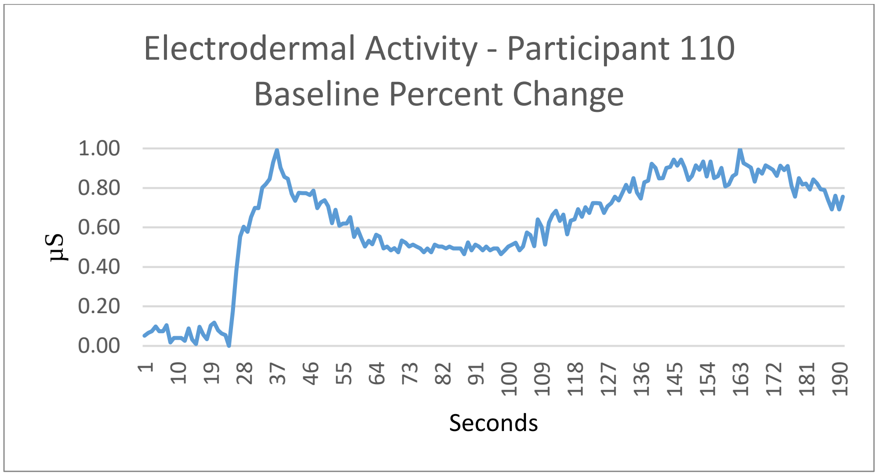

Figure 121 - Participant 110 Electrodermal Activity

Condition 3 – Human-Controlled Avatar

Session 107

The EDA data for participant 107 were shown in Figure 122. There were two distinct spikes in this chart. The first coincided with the start of the session and the second coincided with an especially animated monologue on the part of the participant that started with the words "I'm sorry to hear that happened to you." and proceeded in the most animated engaged sequence of the dialog.

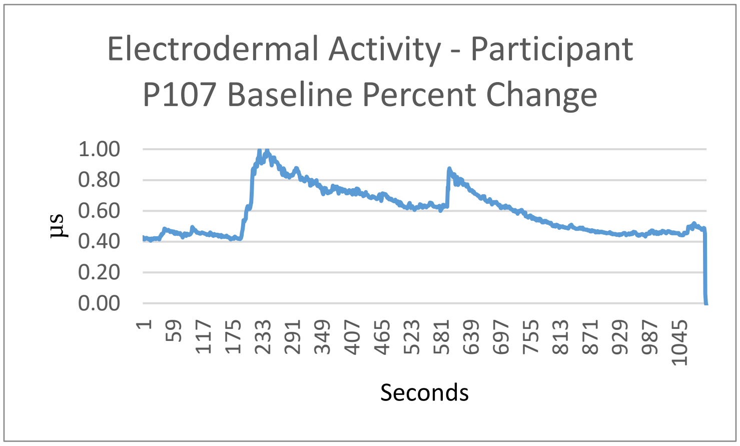

Figure 122 - Participant 107 Electrodermal Activity

Condition 4 – Computer-Controlled Agent

Session 109

In the computer-controlled agent condition, the only participant with EDA data was participant 109, but the data was not characteristic EDA data and was likely noise rather than anything meaningful. The data is shown in Figure 123.

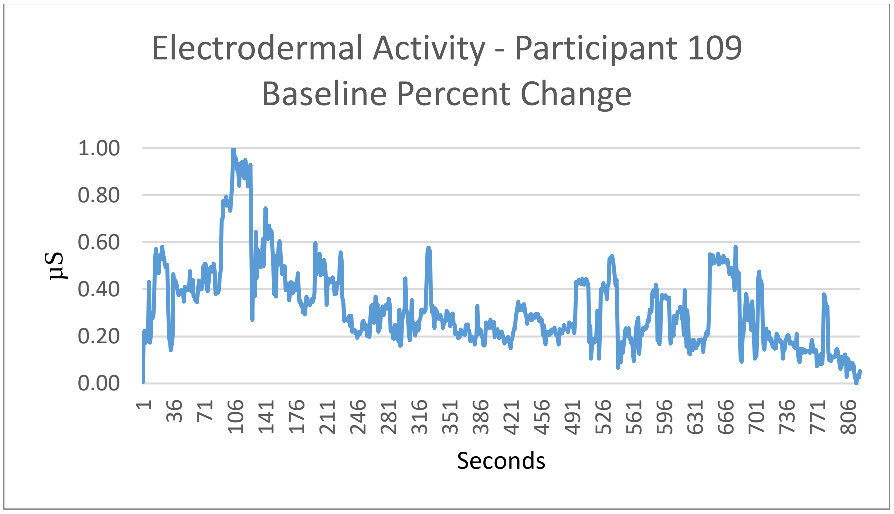

Figure 123 - Participant 109 Electrodermal Activity

Behavioral Indicators

Behavioral indicators, such as hand gestures and nods were tabulated during the dialogs. While research shows that some people will gesture even when they are on a telephone call (Alibali, Heath, & Myers, 2001), gesturing during a dialog might have indicated that the dialog partner had passed the threshold of realism and appeared to have agency. These data were collected to determine if behavioral indicators, such as gestures provided a meaning measure of the sense of agency in a dialog partner. Figure 124 shows the number of gestures per 100 words, which was a standard strategy for gauging conversational gestures (Alibali, Kita, & Young, 2000).

While actor data were collected in the first two conditions, it was not included in the chart for conditions three through five. In several places video footage of the actor (sessions 107, 108, 112, 113, 114, and 115) was unavailable. The video footage for the participant in session 109 was also unavailable due to equipment failure. Condition 4 did not include actor data since it made use of AI.

Setting aside individual differences, it was evident that gesturing behavior was more prevalent in the face-to-face condition (condition 1) and in one session in the avatar condition (condition 3). Gesturing was nearly non-existent when dialoging with the AI character and while texting.

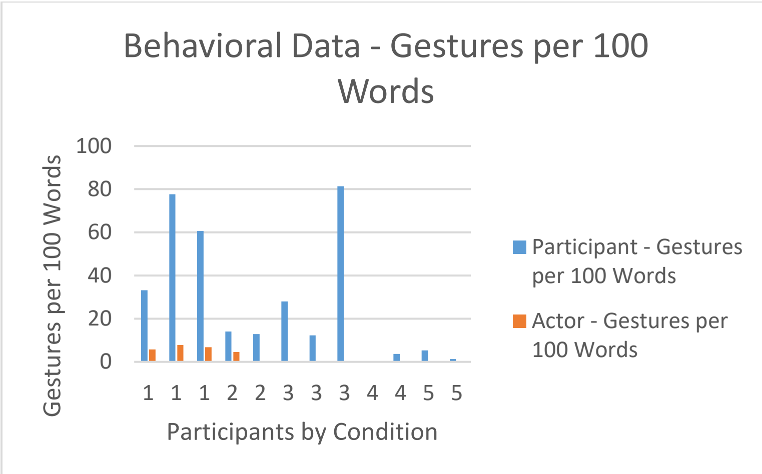

Figure 124 - Gestures per 100 Words for Each Session by Condition

Survey Results

Presence Questionnaire

Figure 125 shows a range of responses that skewed toward the environment being consistent with the real-world. The range of responses for question 1 were 3, 5, 6 and 7 with 6 being extremely consistent. 8.3% of participants selected 3, 33% of participants selected 5, 50% selected 6 and 8.3% selected 7. Participants considered face-to-face and video environment most consistent with the real world followed closely by text. The avatar and agent conditions were perceived slightly less similar to real-world.

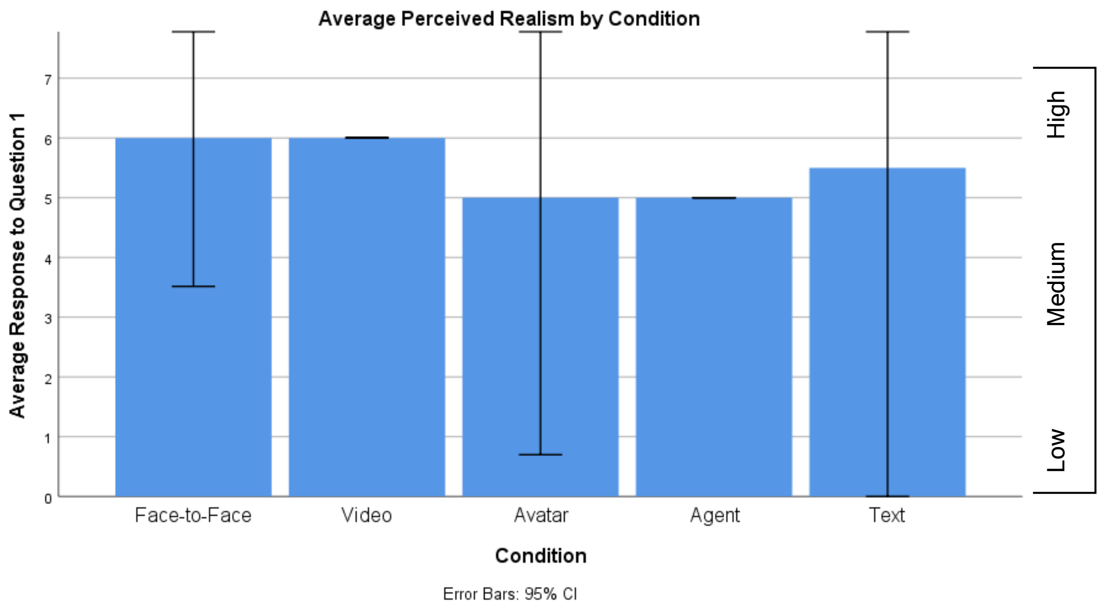

Figure 125 - Average Response to Perceived Realism by Condition

The range of responses for question 2 were 5, 6, and 7 with 7 being extremely consistent. 25% of participants selected 5, 50% selected 6 and 25% selected 7 and indicated responses clustered around the perception of feeling involved in the experience. Figure 126 indicated very little difference between conditions.

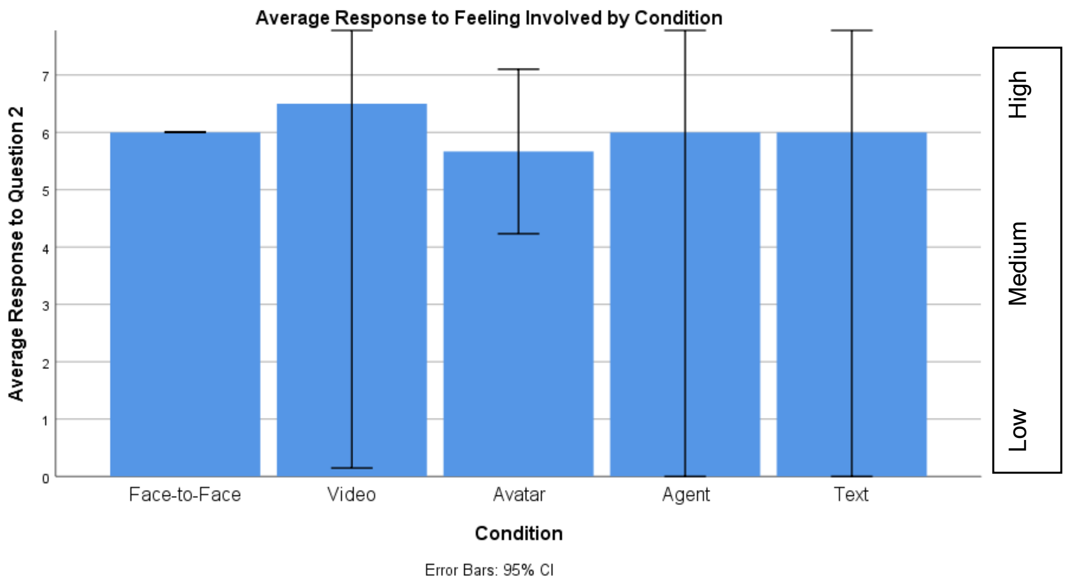

Figure 126 - Average Response to Feeling Involved by Condition

The range of responses for question 3 indicated there was little agreement regarding how much delay was experienced. Higher values indicated longer lag times. 33.3% of participants selected 1, 8.3% of participants selected 2, 16.7% of participants selected 3, 8.3% of participants selected 4, 16.7% selected 5 and 16.7% selected 6. Figure 127 indicates that participants experienced more lag in the text condition followed closely by the agent condition. These were followed by face-to-face then avatar with the least lag in the video condition.

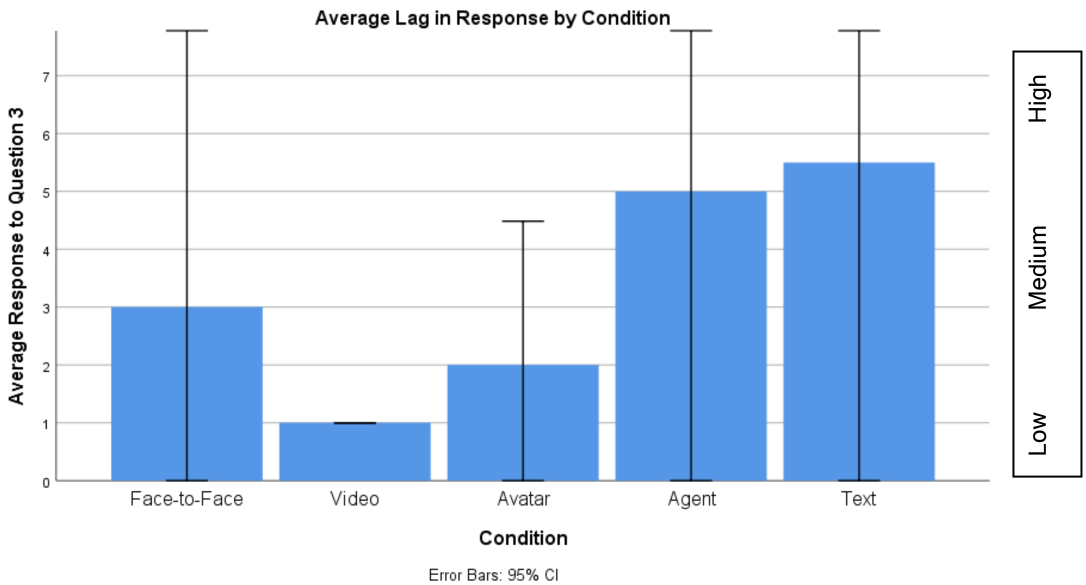

Figure 127 - Average Lag in Response by Condition

Figure 128 indicates participants felt a similar ability to concentrate. 8.3% of participants selected 5, 58.3% of participants selected 6 and 33.3% selected 7.

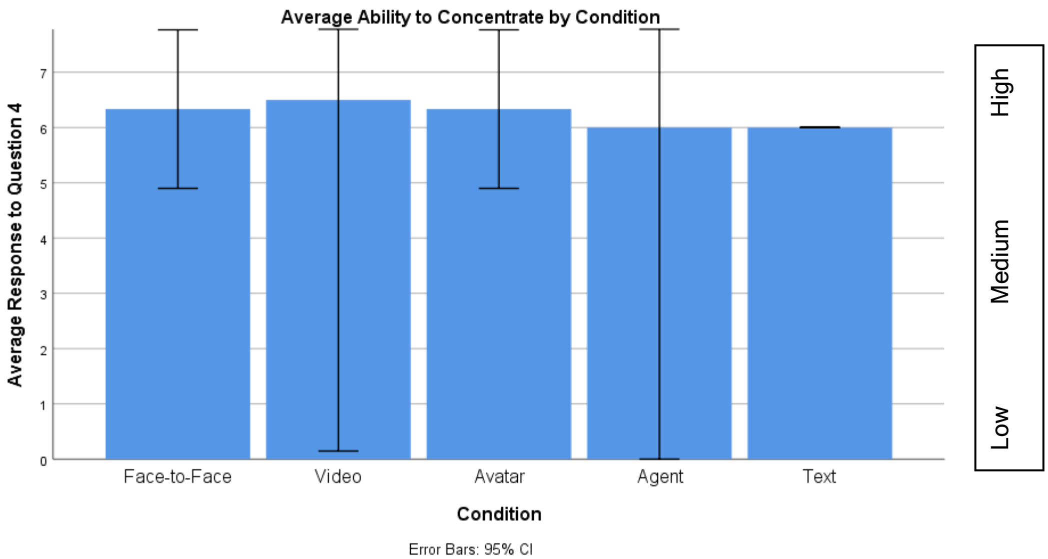

Figure 128 - Average Response to Ability to Concentrate by Condition

Kruskal-Wallis Test was conducted to examine the differences on Presence Questionnaire Results according to the condition. No significant differences (Chi Square=4.667, p=.323, df=4) were found among the five conditions of participants (face-to-face, video, avatar, agent, and text).

Rapport Questionnaire

Question 1 asked about a sense of closeness. Reponses were between 2 and 4 with 8.3% responding with 2, 25% responding with 3 and 66.7% responding with 4. This showed an overall skewing toward closeness or camaraderie between the participants and their interactive partner.

Question 2 of the Rapport Questionnaire asked about figurative distance between the participant and the interaction partner. It was reverse-coded with participants selecting options between 2 and 5 with 8.3% selecting 2, 16.7% selecting 3, 58.3% selecting 4 and 16.7% selecting 4. This indicated a skewing that, in general, showed participants did not feel distance between themselves and the interaction partner. This question was inversely related to question 1.

Question 3 asked about the perceived ability to understand one another. Participants selected options between 2 and 5 with 8.3% responding with 2, 25% responding with 3, 50% responding with 4 and 16.7% responding with 5. This showed an overall skewing toward a sense that the interactive partners understood one another.

Question 4 was reverse-coded and asked if the interactive partner communicated coldness rather than warmth. Participants selected options between 2 and 5 with 8.3% selecting 2, 50% selecting 3, 25% selecting 4 and 16.7% selecting 5. This indicated a very slight skewing that showed participants did not feel a sense of coldness from the interaction partner. This question was the inverse to question five.

Question 5 asked if Jarett communicated warmth. Participants selected options between 2 and 4 with 16.7% selecting 2, 41.7% selecting 3 and 41.7% selecting 4. This indicated a slight skewing that showed participants felt that the interaction partner provided a sense of warmth and caring.

Question 6 had been reverse-coded and asked about the participant maintaining distance in the dialog. Participants selected options between 2 and 5 on question six with 25% responding with 2, 25% responding with 3, 25% responding with 4 and 25% responding with 5. This showed a slight skewing toward a sense that the interactive partners did not maintain a sense of distance.

Question 7 was reverse-paired with question 9. The question asked if the participant felt connected with the interactive partner. Participants selected options between 1 and 4 with 8.3% selecting 1, 16.7% selecting 2, 33.3% selecting 3 and 41.7% selecting 4. This indicated an almost neutral balance with very slight skewing that showed participants felt somewhat close to the interaction partner.

Question 8 was about the sense of respect. Participants selected options between 4 and 5 on question eight with 50% responding with 4 and 50% responding with 5. This showed a significant skewing toward a sense that the interactive partner was respectful.

Question 9 was the reverse-coded counter to question seven. The question asked if the participant felt no connection with the interactive partner. Reverse coded results showed options between 3 and 5 selected with 8.3% selecting 3, 66.7% selecting 4, and 25% selecting 5. This indicated a skewing toward connection.

Question 10 asked if the participant tried to create a sense of closeness or camaraderie. Participants selected options between 2 and 5 on question ten with 16.7% selecting 2, 41.7% selecting 3, 16.7% selecting 4 and 25% selecting 5. This showed a skewing toward a sense of closeness or camaraderie which was associated with question one.

The Rapport questionnaire responses were averaged to provide a number indicating the participant's perception of rapport with the interaction partner. Figure 129 shows the average mean score by condition. Kruskal-Wallis Test was conducted to examine the differences of average results of the Rapport responses based on condition. No significant differences (Chi square = 7.333, p = .158, df=4) were found among the five categories of participants (face-to-face, video, avatar, agent and text). However, the chart does provide a sense of how participants in each category had a similar perception of rapport.

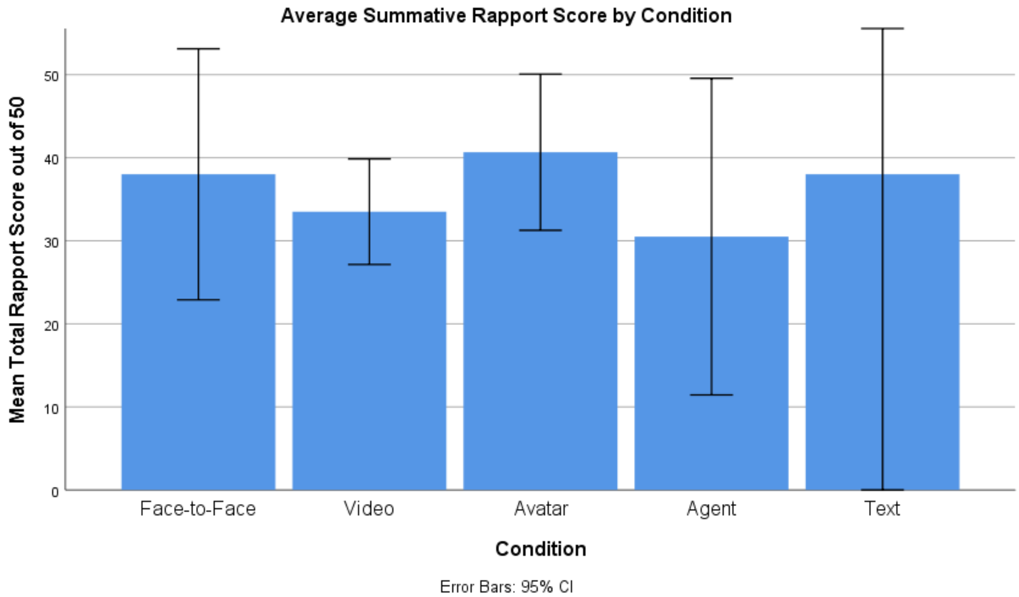

Figure 129 - Mean Rapport Score by Condition

Interaction Questionnaire

Question 1 asked: To what extent were you able to effectively communicate with Jarett? Responses varied from 4 to 7, with 8.3% responding with 4, 41.7% responding with 5, 16.7% responding with 6 and 33.3% responding with 7. This showed an overall skewing of the data to the positive, with people agreeing that they were able to effectively communicate with their interaction partner.

The average participant score by condition is shown in Figure 130. Though not statistically significant, the scores indicated that the avatar condition scored highest, closely followed by video and face-to-face. The agent and text conditions did not appear to provide as effective a strategy to communicate.

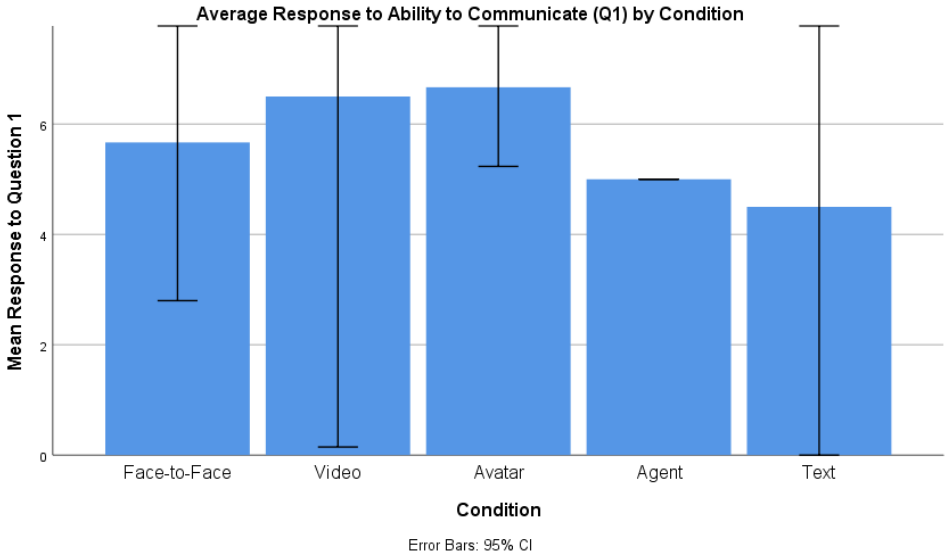

Figure 130 - Average Response to Ability to Communicate by Condition

While the forced-choice question was "To what extent were you able to effectively communicate with Jarett?" the open-ended question was "Why or why not". Responses to the open-ended questions are provided in Table 4.

Table 4 - Open Ended Responses to Question 1 of Interaction Questionnaire

Condition 1 – Face-to-Face	- Didn't know what questions to ask, but he was open to sharing his story - At first, I thought of him as an actor. At the end I believed it was Jarett. I only wanted to hear what he wanted to share.
Condition 2 – Video	- He was very willing to answer the questions I asked of him. - Visually – I could see he was uncomfortable, but tried to keep calm. Eyes were nervous. Verbally – want to tell me what happened in as little words as possible. Very basic description. He seemed like it was just the facts. Didn't want to share much.

Condition 3 – Human-Controlled Avatar	- Able to get story of what happened, its aftermath, and how it made him feel. - Although I felt I was developing a connection with Jarett and developing trust to talk about what happened to him, it was difficult to discern his attitude toward me by looking at the facial expressions of the avatar. - He was able to explain and expound on his particular circumstances.
Condition 4 – AI-controlled Agent	- There was technical difficulty with sound at first. The subject was talking about suicide before I spoke him about his incident. Other questions were answered robotically with a previous statement he had made - Moderately delayed response time, didn't fully understand/properly respond to at least one question.
Condition 5 – Text	- There was not enough time to discuss the event - I felt that he liked my questions and responses.

Question 2 asked: To what extent were you able to understand Jarett? Responses included 3, 4, 6 and 7. 8.3% responded with 3, 8.3% responded with 4, 41.7% responded with 6 and 41.7% responded with 7. This showed a skewing of the data to the positive, with more people agreeing that they were able to understand their interaction partner. The average participant score by condition is shown in Figure 131. Though not statistically significant, the scores indicated that face-to-face, video and agent conditions scored about the same followed by the avatar condition with the text condition scoring lowest.

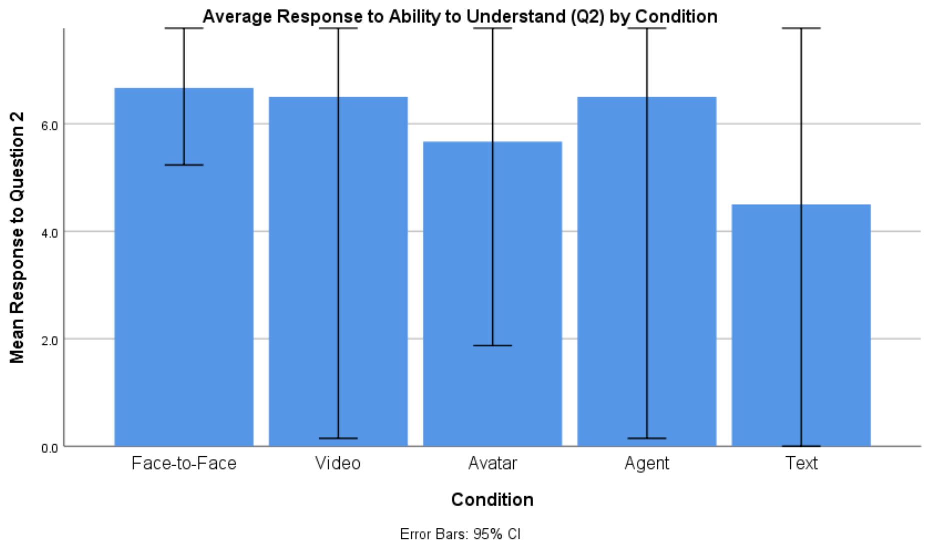

Figure 131 - Average Response to Ability to Understand by Condition

While the forced-choice question was “To what extent were you able to understand Jarett?” the open-ended question was “If you couldn’t understand him, what was the barrier to understanding?” Responses to the open-ended questions are provided in Table 5.

Table 5 - Open Ended Responses to Question 2 of Interaction Questionnaire

Condition 1 – Face-to-Face	- There were no issues understanding Jarett
Condition 2 – Video	- There were no issues understanding Jarett.
Condition 3 – Human-Controlled Avatar	- Difficulty deciphering his emotions through body language of avatar. - Jarett sounded more engaged than his avatar suggested. - Slight noise interference (mic garble/echo).

Condition 4 – AI-controlled Agent	- I was a little distracted by his body language. He kept fidgeting in his chair with his hand.
Condition 5 – Text	- Not sure, but maybe some miscommunication in asking who did what.

Question 3 asked: To what extent were you frustrated during the dialog? Responses included 1, 2, 4, 5 and 6. 41.7% responded with 1, 16.7% responded with 2, 16.7% responded with 4, 8.3% responded with 5 and 16.7% responded with 6. This showed a skewing of the data to indicate limited frustration. The average participant score by condition for question 3 is shown in Figure 132. Though not statistically significant, the scores indicated that the agent condition scored highest in frustration. Avatar and face-to-face were lower but about the equal to one another followed by video and text appeared to have had the least amount of frustration.

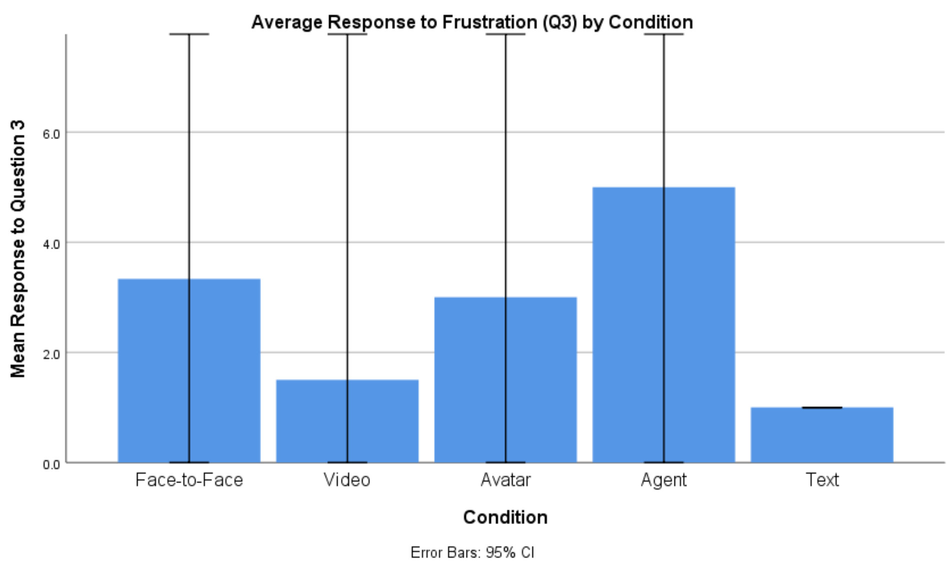

Figure 132 - Average Response to Frustration by Condition

While the forced-choice question was "To what extent were you frustrated during the dialog?" the open-ended question was "If you experienced frustration, what was the source?" Responses to the open-ended questions are provided in Table 6.

Table 6 - Open Ended Responses to Question 3 of Interaction Questionnaire

Condition 1 – Face-to-Face	- Thinking of what to ask, how I was supposed to get information from him. - Felt a little frustrated that he was discharged for something not his fault.
Condition 2 – Video	- No sources of frustration.
Condition 3 – Human-Controlled Avatar	- Thinking of what to ask, how I was supposed to get information from him. - Jarett was discharged but assailants weren't.
Condition 4 – AI-controlled Agent	- The subject did not answer my direct question. Instead he repeated verbatim the previous statement about this incident. - Didn't allow me to break the ice, he basically just jumped into what happened.
Condition 5 – Text	- Thinking of what to ask, how I was supposed to get information from him.

Question 4 asked: To what extent did you feel understood? Responses were 2 between and 7. 8.3% responded with 2, 16.7% responded with 3, 2, 8.3% responded with 4, 16.7% responded with 5, 16.7% responded with 6 and 33.3% responded with 7. This showed a wide range of responses which are shown in Figure 133. Though not statistically significant, the scores indicated that participants in the video condition showed greater understanding followed by face-to-face, avatar then the text condition. The agent condition rated least understood.

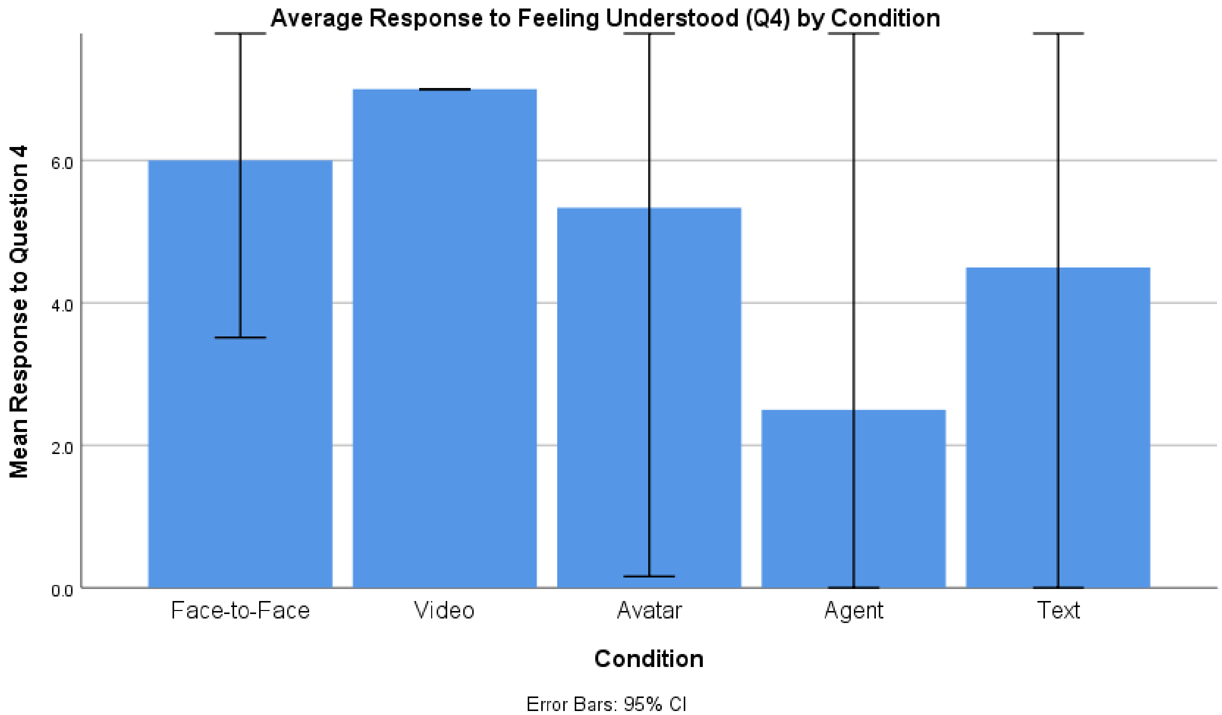

Figure 133 - Average Response to Feeling Understood by Condition

While the forced-choice question was "To what extent did you feel understood?" the open-ended question was "What indication did Jarett give that he did or did not understand you?" Responses to the open-ended questions are provided in Table 7.

Table 7 - Open Ended Responses to Question 4 of Interaction Questionnaire

Condition 1 – Face-to-Face	- I didn't really say much to be understood. - I think we understood each other pretty well.

Condition 2 – Video	- I think he understood what was asked.
Condition 3 – Human-Controlled Avatar	- He answered all my questions completely - I had no visual of how I may have appeared to Jarett. Had it been a regular webcam where I could see myself, I might have made adjustments in my body language to better match the scenario. - I don't think he didn't understand, I think I phrased my question incorrectly to get the answer I was looking for. I should have been clearer.
Condition 4 – AI-controlled Agent	- He did not answer my direct questions. - I asked him to tell me a little bit about himself, and he didn't give me the opportunity to introduce myself or talk a little about myself.
Condition 5 – Text	- Jarett was interested in my suggestions.

Question 5 asked: To what extent did you notice emotional indicators in Jarett? Responses were between 2, 4, 5, 6 and 7. 16.7% responded with 2, 16.7% responded with 4, 25% responded with 5, 16.7% responded with 6 and 25%% responded with 6 and 33.3% responded with 7. This showed a heavy weight toward participants being able to perceive emotional indicators. The average participant score by condition for question 5 is shown in Figure 134. Though not statistically significant, the scores in the face-to-face condition were the highest possible followed by the video and agent conditions, followed by the avatar condition and text provided the least emotional indicators.

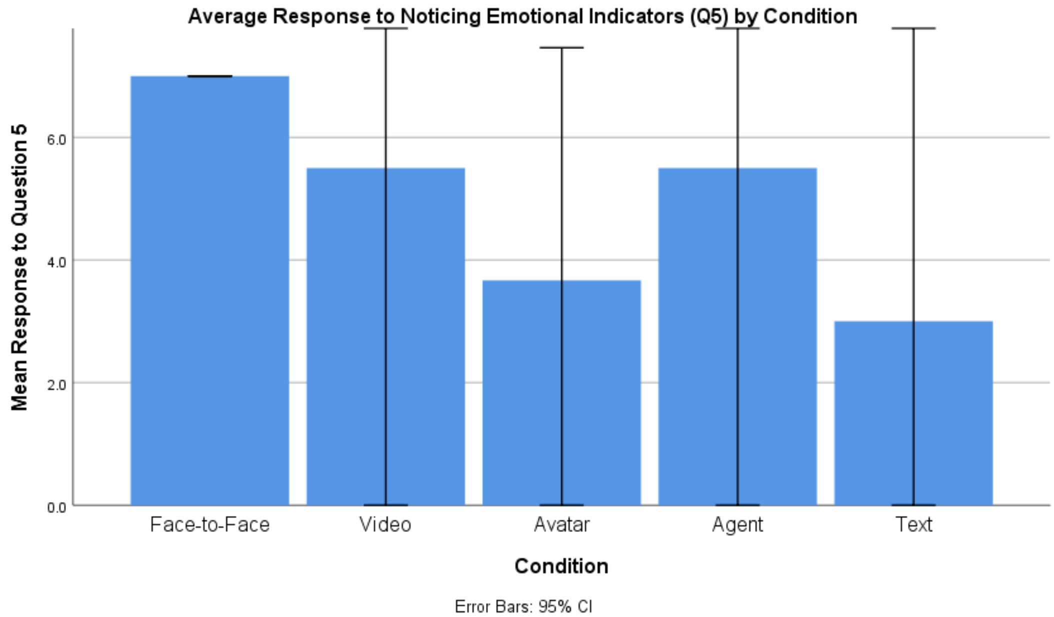

Figure 134 - Average Response to Noticing Emotional Indicators by Condition

While the forced-choice question was "To what extent did you notice emotional indicators in Jarett?" the open-ended question was "What were the emotional indicators you noticed?" Responses to the open-ended questions are provided in Table 8.

Table 8 - Open Ended Responses to Question 5 of Interaction Questionnaire

Condition 1 – Face-to-Face	- He seemed shaky in his hands and arms and in his voice. He maintained eye contact though. - Visibly shaking. - He seemed sad about it, and I felt he still struggled with it.
Condition 2 – Video	- I could tell that discussing certain aspects of the events to be disturbing to him. - Eye movement. Not a lot of head movement. Didn't notice posture.

Condition 3 – Human-Controlled Avatar	- Changes in tone - Visual and vocal indicators were often contradictory. Ex: hearing him laugh but only seeing a faint smile with “cold eyes” on the avatar. - Sighs/heavy breath, pausing.
Condition 4 – AI-controlled Agent	- Fidgeting, using hand movements. Leg movements. - Shaky voice, constant fidgeting.
Condition 5 – Text	- There were slow responses which could indicate that he was upset, but could just be slow at typing or not fully paying attention. - Jarett seemed relieved to be out of the job where he was assaulted.

Question 6 asked: To what extent did you trust Jarett? Responses were between 4, 5, 6 and 7. 25% responded with 4, 8.3% responded with 5, 33.3% responded with 6, and 33.3% responded with 7. This showed a heavy weight toward participants trusting their interaction partner. The average participant score by condition for question 6 is shown in Figure 135. Though not statistically significant, the scores in the face-to-face condition are the highest followed by avatar then video with agent and text scoring lowest.

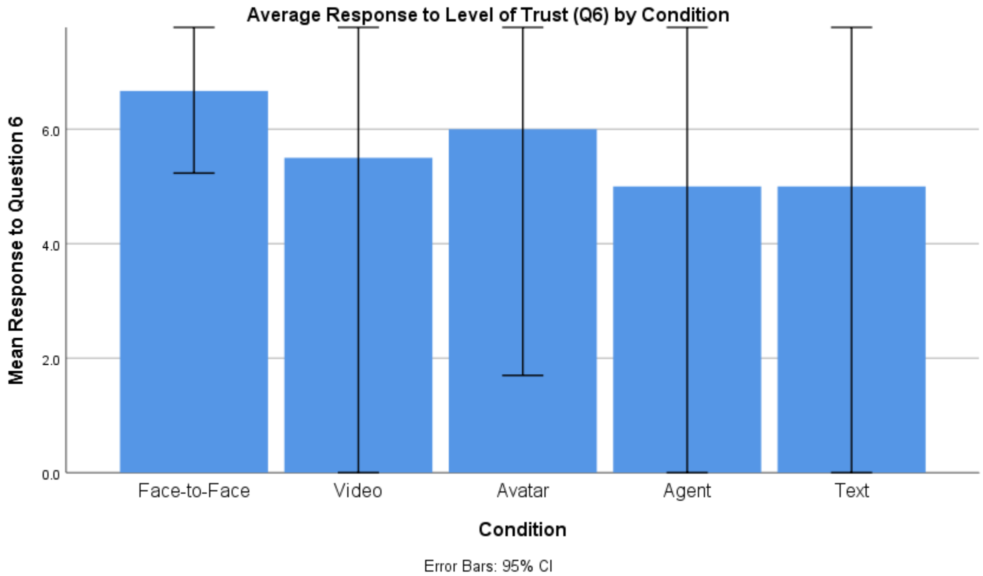

Figure 135 - Average Response to Level of Trust by Condition

While the forced-choice question was "To what extent did you trust Jarett?" the open-ended question was "Why or why not?" Responses to the open-ended questions are provided in Table 9.

Table 9 - Open Ended Responses to Question 6 of Interaction Questionnaire

Condition 1 – Face-to-Face	- He seemed vulnerable - He was very open and willing to share something traumatic that had happened to him.
Condition 2 – Video	- This was an initial encounter. There would need to be many more sessions to validate the events as he describes them versus his possible embellishment. - Details were uncomfortable to share. Seemed closed. He wanted you to know he was uncomfortable with what he experienced, but wanted you to know that it happened.

Condition 3 – Human-Controlled Avatar	- Have no reason to not trust. - Innocent until proven guilty.
Condition 4 – AI-controlled Agent	- I felt it was animation and robotic. But wasn't sure and didn't want to be disrespectful. - Body language.
Condition 5 – Text	- It never went one way or the other. - Jarett seemed real and was handling a difficult situation.

Question 7 asked: To what extent did you experience stress during the interchange? Responses were between 1 through 5 and 7. 8.3% responded with 1, 33.3% responded with 2, 16.7% responded with 3. 8.3% responded with 4, 16.7% responded with 5, and 16.7% responded with 7. This showed a wide gap in responses regarding participant's perception of stress during the interaction. The average participant score by condition for question 7 is shown in Figure 136. Though not statistically significant, the scores in the agent condition are highest, followed by the face-to-face condition, then video and avatar conditions. Text scores were lowest for stress.

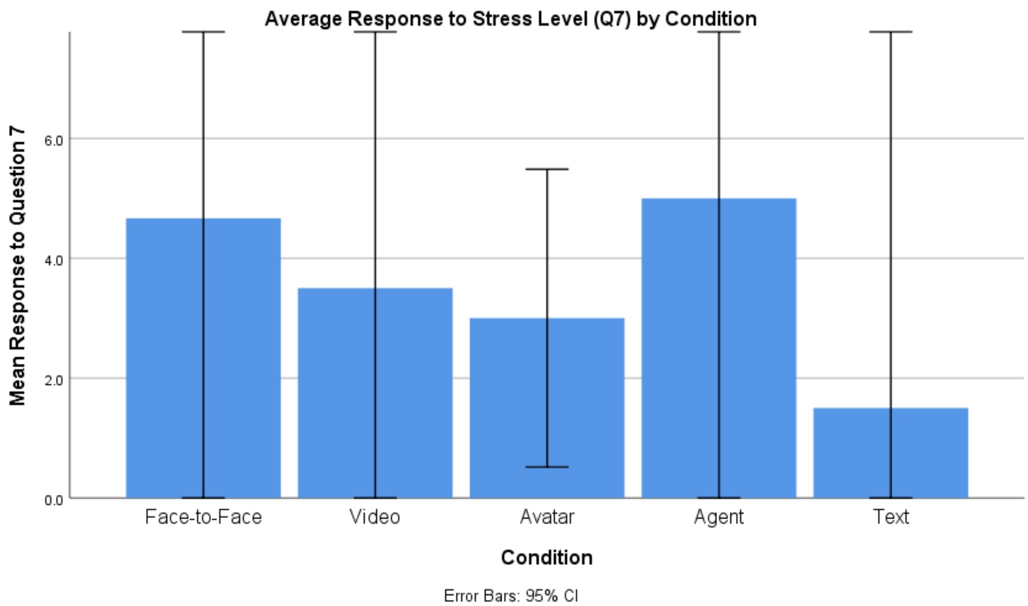

Figure 136 - Average Response to Stress Level by Condition

While the forced-choice question was "To what extent did you experience stress?" the open-ended question was "If so, what was the source of the stress?" Responses to the open-ended questions are provided in Table 10.

Table 10 - Open Ended Responses to Question 7 of Interaction Questionnaire

Condition 1 – Face-to-Face	- Empathizing with Jarett's story - I felt a little stressed knowing that he was about to talk about something bad that had happened to him.
Condition 2 – Video	- Uncomfortable with asking him questions. Didn't know what to ask.
Condition 3 – Human-	- To become comfortable enough with a stranger. To ask the hard questions. - Asking personal questions to someone I don't know or just met.

Controlled Avatar	
Condition 4 – AI-controlled Agent	- The act of his incident was disturbing. - Hearing descriptions of what happened.
Condition 5 – Text	- No stress, so much as wanting to make sure not to cause any.

Question 8 asked: To what extent was the interaction natural and believable? Responses were between 1 through 5 and 7. 8.3% responded with 1, 33.3% responded with 2, 16.7% responded with 3. 8.3% responded with 4, 16.7% responded with 5, and 16.7% responded with 7. This showed a wide range in responses regarding participant's perception of the interaction. The average participant score by condition for question 8 is shown in Figure 137. Face-to-face and video scored highest with avatar and text slightly lower followed by agent as the lowest scoring condition.

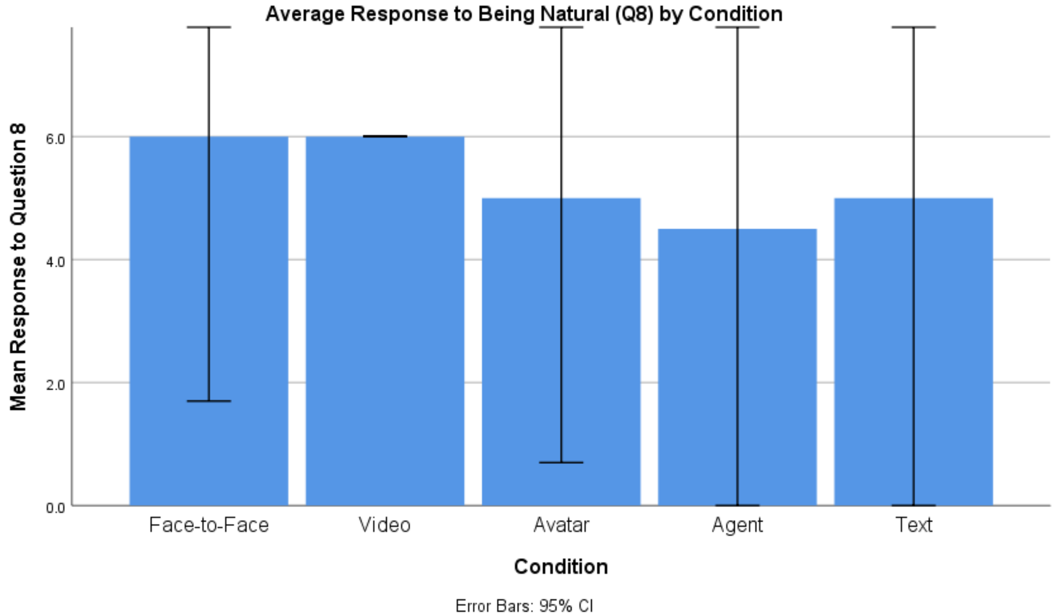

Figure 137 - Average Response to Being Natural by Condition

While the forced-choice question was "To what extent was the interaction natural and believable?" the open-ended question was "If the interaction was not natural, what made it feel unnatural?" Responses to the open-ended questions are provided in Table 11.

Table 11 - Open Ended Responses to Question 8 of Interaction Questionnaire

Condition 1 – Face-to-Face	- Probably the study setup. The preparation beforehand skewed my perception of the scenario. - It felt very natural!
Condition 2 – Video	- Once you get focused on Jarett rather than the equipment it gets more believable.

Condition 3 – Human-Controlled Avatar	- Differences in vocal and visual cues of emotion. - My own self-doubt on new interactions-I felt clumsy.
Condition 4 – AI-controlled Agent	- The delay in response and that he did not answer my direct questions. - Delayed response time, “fading” from one appearance to another (e.g., middle of speaking, then suddenly stopping and resuming fidgeting). He also didn’t ask me my name, or anything about me.
Condition 5 – Text	- No responses

Question 9 asked: To what extent was Jarett trustworthy? Responses were between 4 through 7. 25% responded with 4, 16.7% responded with 5, 33.3% responded with 6. And 25% responded with 7. There was heavy leaning, in the responses, toward the participants having the sense that Jarett was trustworthy. The average participant score by condition for question 9 is shown in Figure 138. The scores for face-to-face were highest followed closely by video, then avatar, then text and finally agent was rated lowest.

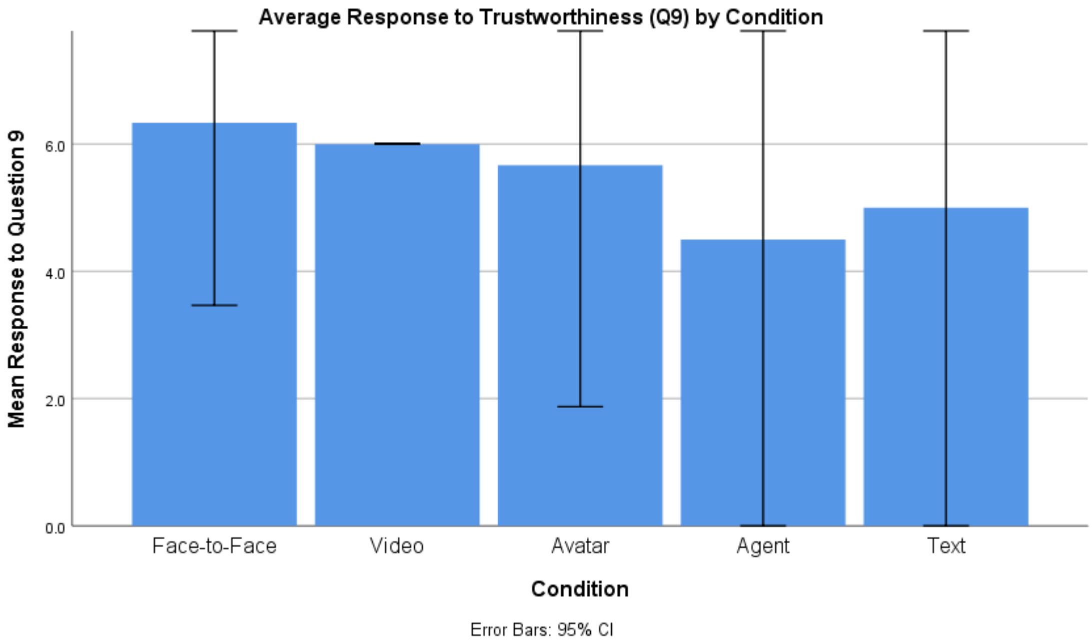

Figure 138 - Average Response to Trustworthiness by Category

While the forced-choice question was "To what extent was Jarett trustworthy?" the open-ended question was "What made Jarett seem trustworthy or untrustworthy?" Responses to the open-ended questions are provided in Table 12.

Table 12 - Open Ended Responses to Question 9 of Interaction Questionnaire

Condition 1 – Face-to-Face	- Just him being open to share his story. - His raw emotional response made his testimony trustworthy. - His openness and willing to share his experience made him trustworthy to me.
Condition 2 – Video	- This was an initial encounter. There would need to be many more sessions to validate the events as he describes them versus his possible

	embellishment…he seems to be believable, but more discussions would be needed with more than just him. - Trustworthy because of his measured detail and response. He seemed uncomfortable and that is how I would expect a person to act.
Condition 3 – Human-Controlled Avatar	- No reason to not trust him. - Avatar appeared cold and callus. - He offered more detail than I asked ->back-story into event.
Condition 4 – AI-controlled Agent	- I felt it may have been robotic but wasn't sure. Not that I didn't trust him – more that I wasn't trusting the system or the situation. - Could not shake his hand, and he didn't bother to ask my name.
Condition 5 – Text	- Not enough information to decide. - Jarett had a real problem and was serious about getting help.

Question 10 asked: To what extent was Jarett persuasive? There was a wide range of responses. 25% responded with 1, 8.3% responded with 2, 8.3% responded with 3, 16.7% responded with 4, 16.7% responded with 5, 16.7% responded with 6 and 8.3% responded with 7. The average participant score by condition for question 10 is shown in Figure 139. The average scores suggested that even though participants interpreted the meaning of this question differently, the scores in the face-to-face and text were highest with avatar being lowest.

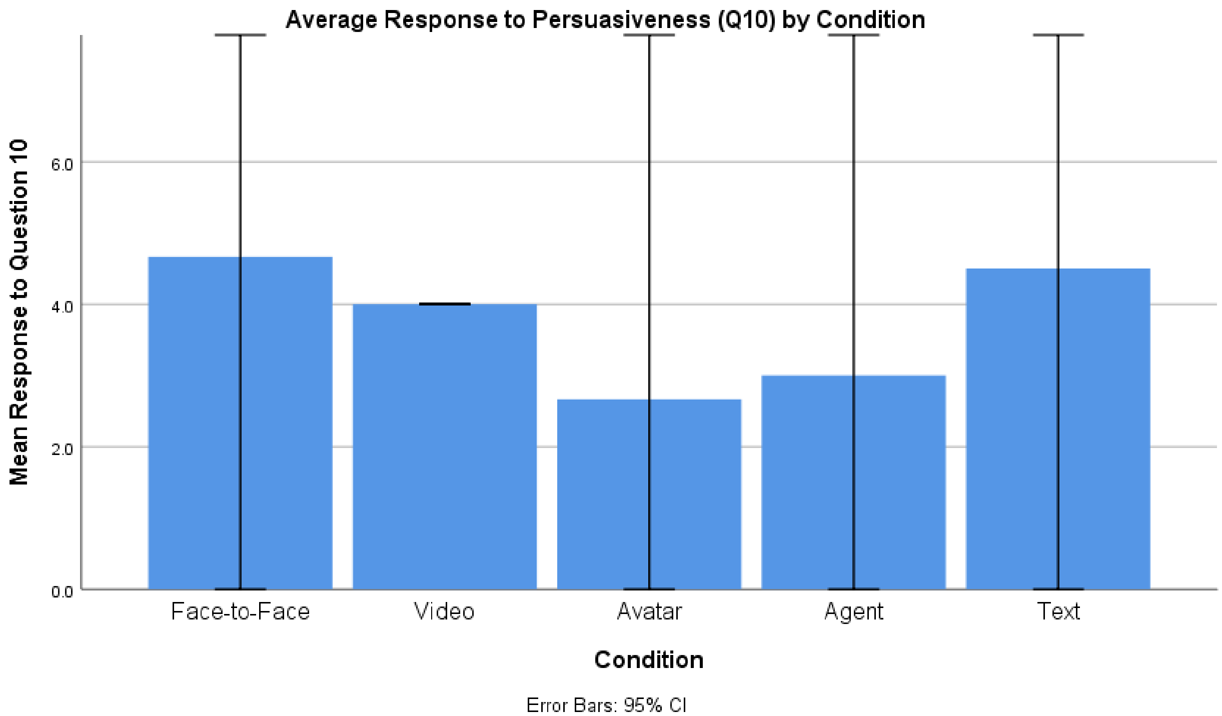

Figure 139 - Average Response to Persuasiveness by Condition

While the forced-choice question was "To what extent was Jarett persuasive?" the open-ended question was "What made Jarett seem persuasive or unpersuasive?" Responses to the open-ended questions are provided in Table 13.

Table 13 - Open Ended Responses to Question 10 of Interaction Questionnaire

Condition 1 – Face-to-Face	- I didn't take it as he was trying to persuade me, I was just listening. - His details and ability to tell his story. - He was neither persuasive nor unpersuasive.
Condition 2 – Video	- This was an initial encounter. There would need to be many more sessions to validate the events as he describes them versus his possible

	embellishment…he seems to be believable, but more discussions would be needed with more than just him. - - Unpersuasive because he just gave minimal facts. He didn't try to prove he was assaulted.
Condition 3 – Human-Controlled Avatar	- He didn't persuade me of anything. - Not sure he was persuading me or attempting to. - It was "matter of fact".
Condition 4 – AI-controlled Agent	- Our initial hello, when I asked if he could hear me. The rest seemed ok except when he repeated verbatim the attack – when I questioned the suicidal thoughts. - He stuck to the facts; I didn't get the sense that he was trying to persuade me of anything.
Condition 5 – Text	- Not sure that anything was said that "persuasive" would make sense as a description. - Jarett persuaded me that he felt good about getting therapy.

Question 11 asked: To what extent was Jarett's voice natural, if you heard him? Responses ranged across the entire scale. 25% responded with 1, 8.3% responded with 2, 8.3% responded with 3, 16.7% responded with 4, 16.7% responded with 5, 16.7% responded with 6 and 8.3% responded with 7. Answers indicated that participants interpreted this question in very different ways. The average participant score by condition for question 11 is shown in Figure 140. The video condition scored highest followed closely by the avatar condition. These were followed by face-to-face then agent. This question did not apply to text.

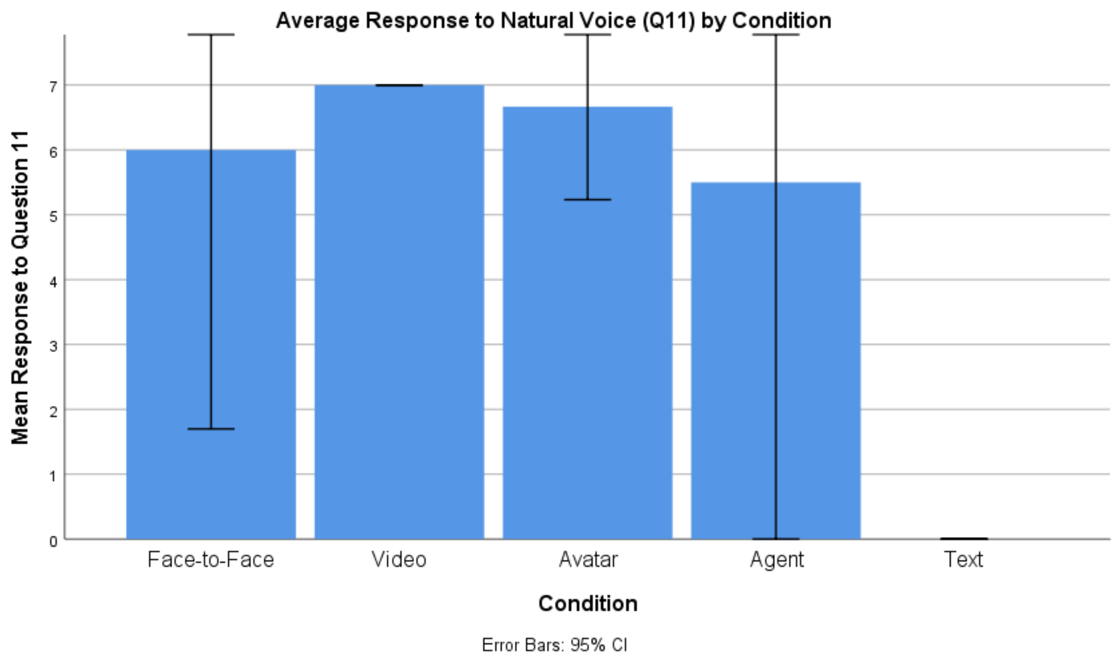

Figure 140 - Average Response to Natural Voice by Condition

While the forced-choice question was "To what extent was Jarett's voice natural, if you heard him?" the open-ended question was "What made the voice seem natural or unnatural?" Responses to the open-ended questions are provided in Table 14.

Table 14 - Open Ended Responses to Question 11 of Interaction Questionnaire

Condition 1 – Face-to-Face	- It seemed a little dramatic/acting. - He sounded like any normal guy. - He was talking in person at me.

Condition 2 – Video	- He didn't seem like he was reading from a script. Used "Um, Ah, etc." Responses seemed measured. Little too quick to respond. I would need more time to answer.
Condition 3 – Human-Controlled Avatar	- How it became somewhat difficult for him to talk about his experiences. - Slight echo/mic noise but real sounding.
Condition 4 – AI-controlled Agent	- I thought I was speaking to an actual person to begin with so I had no reason not to believe it would be unnatural. If it was simulated it was good. - Video recording of a real person, at least it appeared so.
Condition 5 – Text	- This question was not applicable.

Question 12 asked: To what extent did Jarett's movement seem natural, if you could see him? Responses were 1, 4, 5, 6 and 7. 16.7% responded with 1, 25% responded with 4, 8.3% responded with 5, 25% responded with 6 and 25% responded with 7. The average participant score by condition for question 12 is shown in Figure 141. The agent condition scored highest followed by video then face-to-face, followed at a distance by the avatar condition. This question did not apply to text.

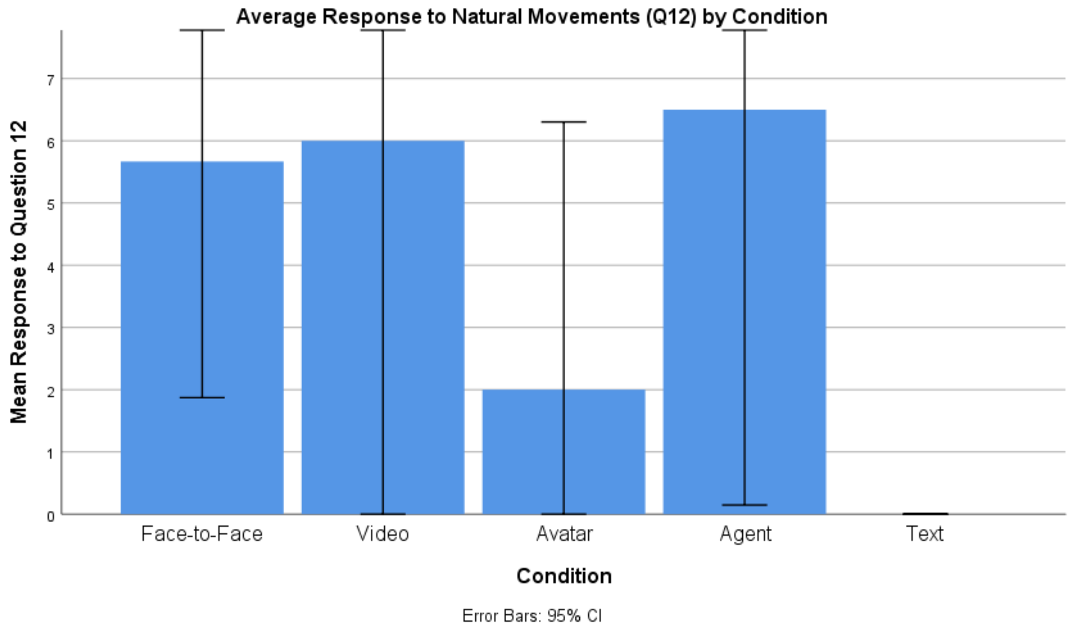

Figure 141 - Average Response to Natural Movements by Condition

While the forced-choice question was “To what extent did Jarett’s movement seem natural, if you could see him?” the open-ended question was “What made the movement seem natural or unnatural?” Responses to the open-ended questions are provided in Table 15.

Table 15 - Open Ended Responses to Question 12 of Interaction Questionnaire

Condition 1 – Face-to-Face	- Though shaking, I feel like, given what he spoke about, his reaction and movements seemed completely natural. - His movement was very natural, and very normal.
Condition 2 – Video	- I was focused on his eyes. Didn’t pay attention to body. His wave was unnatural.

Condition 3 – Human-Controlled Avatar	- Arms were just straight/sometimes twitching. Only mouth movement. - Shoulders never moved with arms, no movement above mouth (crinkling eyes, moving cheeks). - Natural facial movements & hand “ticks” but victims generally don’t sit with arms straight down by their sides. The stillness of the rest of his body is unnatural.
Condition 4 – AI-controlled Agent	- A person who’d been through what he had been through would be fidgety and have lots of movement – in my opinion. - Typical indicators of nervousness while discussing an uncomfortable topic.
Condition 5 – Text	- This question was not applicable.

Question 13 asked: To what extent did you like Jarett? Responses were from 4 through 7. 33.3% responded with 4, 16.7% responded with 5, 33.3% responded with 6 and 16.7% responded with 7. These results suggested a skewing from neutral to very positive indicating friendly feelings toward Jarett. The average participant score by condition for question 13 is shown in Figure 142. Scores were generally high with face-to-face highest and the avatar and text conditions being slightly lower than the rest.

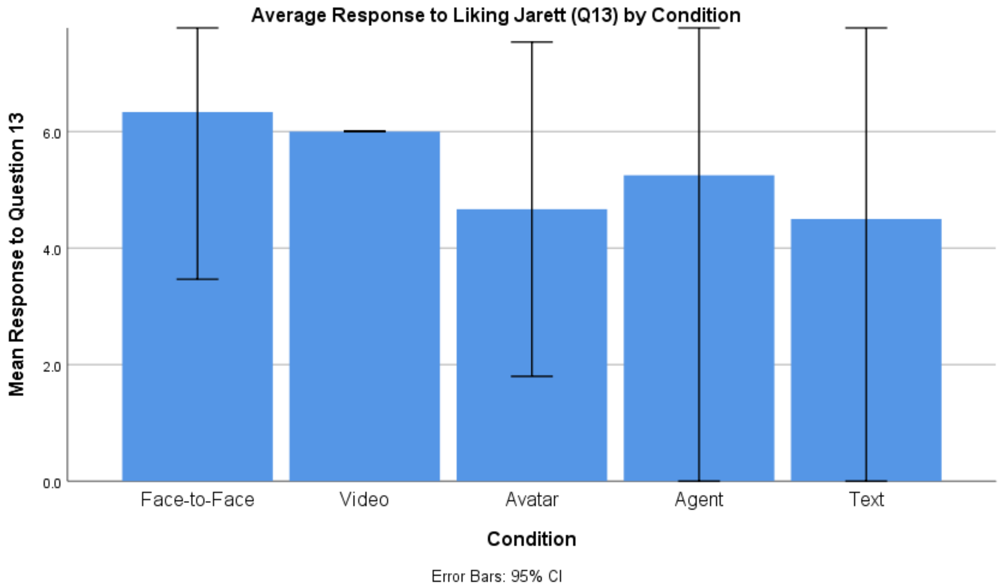

Figure 142 - Average Response to Liking Jarett by Condition

Question 14 asked: To what extent did you dislike Jarett? Responses were 1, 2 and 4. 41.7% responded with 1, 33.3% responded with 2, and 25% responded with 4. These results were a pretty clear inverse of the previous question with responses ranging from neutral to negative, supporting the idea that participants had a generally friendly feeling toward Jarett. The average participant score by condition for question 14 is shown in Figure 143. Scores were generally low with the avatar, agent and text conditions being slightly higher (neutral) than face-to-face and video.

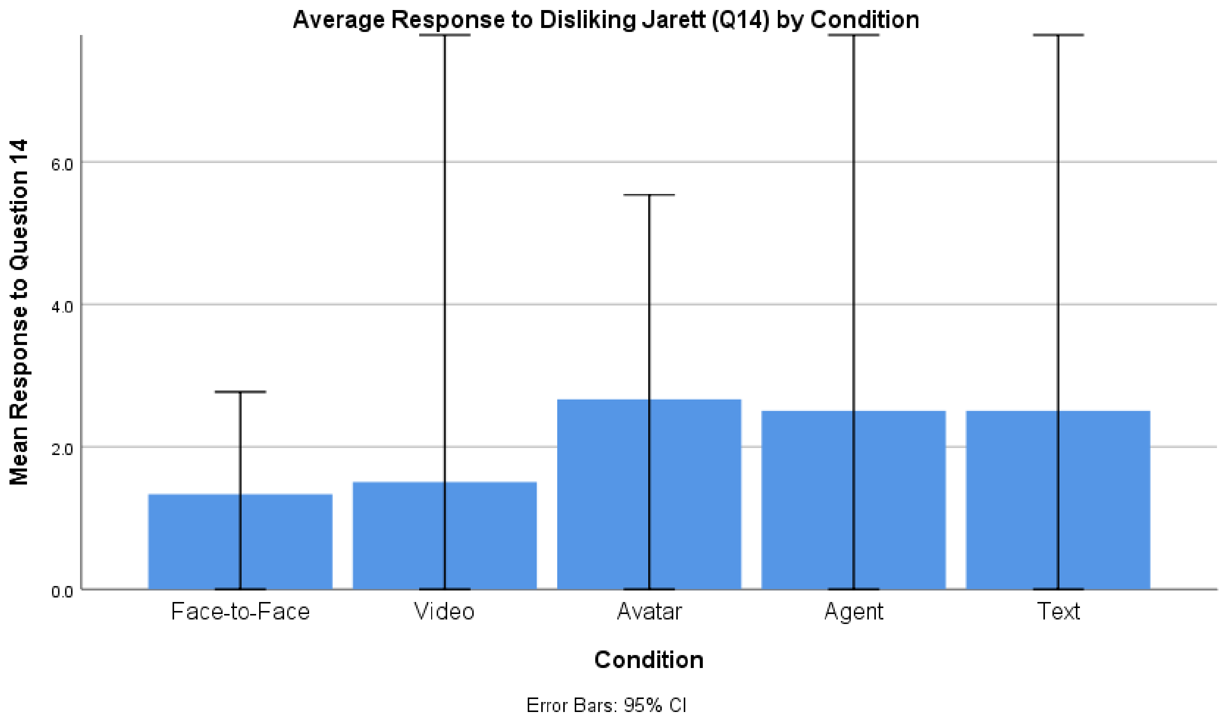

Figure 143 - Average Response to Disliking Jarett by Condition

While the forced-choice questions were "To what extent did you like Jarett?" and "To what extent did you dislike Jarett" the open-ended question was "What did you like or dislike about Jarett?" Responses to the open-ended questions are provided in Table 16.

Table 16 - Open Ended Responses to Question 13/14 of Interaction Questionnaire

Condition 1 – Face-to-Face	- His tone was respectful and open to conversation. - I liked how honest he was being and his kindness. - Didn't dislike anything. Liked that he was able to share what happened and trusted me enough to tell me.
Condition 2 – Video	- Seemed like a nice guy. - He looked "relatable".
Condition 3 – Human-Controlled Avatar	- Liked: openness perceived in voice (not really the avatar). - Liked his openness and storytelling.
Condition 4 – AI-controlled Agent	- I was brought here to help assess a situation – there was no reason for me not to like him. I felt compassion, disgust, and emotional for his situation. - Never had a connection, I mostly just listened to him.
Condition 5 – Text	- Again, not enough information.

Question 15 asked: To what extent did you enjoy the interchange? Responses were 1 through 6. 16.7% responded with 1, 8.3% responded with 2, 8.3% responded with 3, 33.3% responded with 4, 8.3% responded with 5 and 25% responded with 6. These results suggested disagreement among the participants in the meaning of the question. The average participant score by condition for question 15 is shown in Figure 144. Scores were all over with video being high followed by text then face-to-face. The avatar and agent scores were low. Participants viewed this question in different ways based on the dialog topic and the actual interchange.

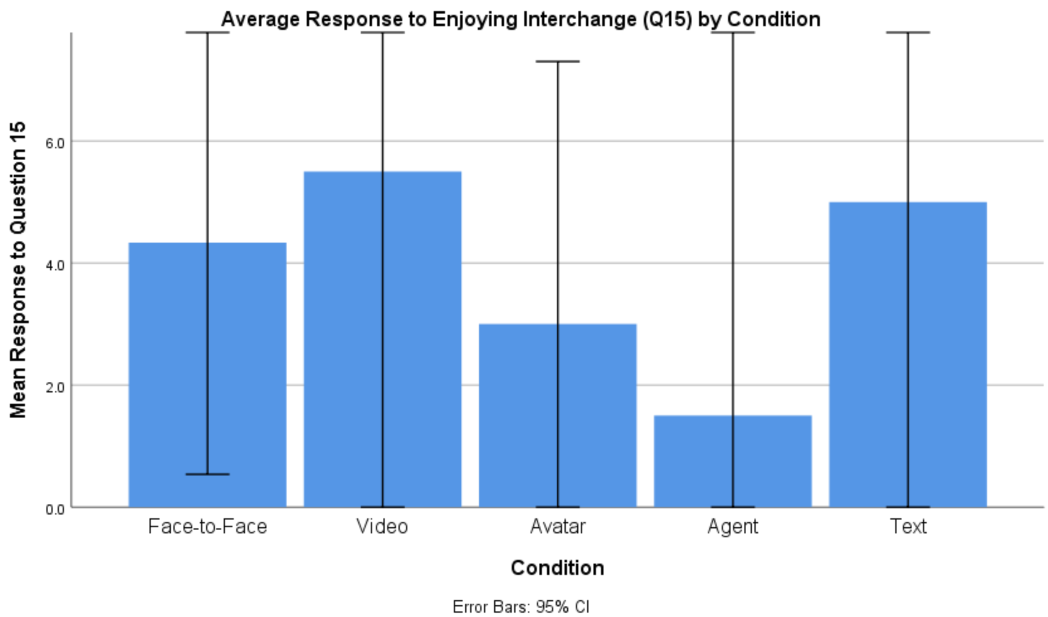

Figure 144 - Average Response to Enjoying the Interchange by Category

While the forced-choice questions were "To what extent did you enjoy the interchange?" the open-ended question was "What did you find enjoyable or unenjoyable about the interchange?" The responses to the open-ended questions are provided in Table 17.

Table 17 - Open Ended Responses to Question 15 of Interaction Questionnaire

Condition 1 – Face-to-Face	- It wasn't an "enjoyable" subject, but I think I did learn from Jarett. - Jarett was easy to speak with, and the unenjoyable part was the topic.

	- Jarett seemed like a nice person. It was nice chatting with him, but didn't find what happened to him enjoyable. I think listening to people go through tough experiences is difficult.
Condition 2 – Video	- The equipment was a bit of a bother, but overall, the experience was fine. - Neutral: It's a difficult subject to discuss when you haven't had the experience and are unfamiliar with the individual.
Condition 3 – Human-Controlled Avatar	- Nothing to really enjoy about talking with someone about their assault. - Disliked: Knowing he could see me, but I could not see him nor could I see myself. - Enjoyable – the tech in this is so cool! Unenjoyable – the topic & my role.
Condition 4 – AI-controlled Agent	- I didn't like the idea that this could happen to someone in the service. I was drawn in to the ordeal & would love to help but I wouldn't say ANY of it was "enjoyable." - Uncomfortable subject of discussion, plus nervousness.
Condition 5 – Text	- It was good to hear that he had made progress in processing the event. - I wanted to help Jarett and enjoyed offering suggestions of help.

Social Phobia Inventory (SPIN)

The Social Phobia Inventory separated people into two categories, those with social phobia and those without. The survey contained 17 questions and was scored on a Likert-type scale with 0 being "Not at All" and 4 being "Extremely". Summative scores greater than 19 indicated participants who experience social anxiety. Three individuals scored above the threshold score for social anxiety. Two individuals were in condition 1 had scores above 19 (36 and 32). One individual was in condition 4 and also had a score above 19 (25).

CHAPTER FIVE: DISCUSSION

Objectives

This study was a novel approach to explore measurement strategies that would indicate that an individual believes his/her interaction partner had agency. Three different measurement strategies, biometric, behavioral, and survey responses were explored to inform future hypotheses on interpersonal interactions in an M&S environment. Several factors influenced the ability to gain insight based on these measures.

Low Number of Participants

One important challenge in this study was scheduling sessions. There were four critical attendees: the actor, the participant, the proctor, and the EEG technician. In nearly every case each of these people had a full-time job along with other obligations, such as family matters and health issues that made it difficult to schedule sessions.

Once schedules aligned enough to schedule sessions, the unpredictability of the time it would take to prepare participants made it impossible to schedule more than two participants per day. It was challenging to get every sensor to activate on the 24 channel EEG. The rate of participant data collection shown in Figure 10 graphically shows the low number of participants. In order to run this experiment in the future, more reliable equipment and a dedicated staff is recommended.

Equipment

The EEG equipment took over an hour to set up. This process could have affected the results of the data by fatiguing the participants and putting them in a less-than-receptive state of mind.

The EDA equipment was discontinued prior to its use in this study. Customer support and analysis software were unavailable. The systems did not consistently turn on at the start of the session nor turn off when finished. Further exploration would need to be conducted to find reliable and cost-effective strategies to collect EDA data for it to be used in future research.

Time to Interact

The 8 to 10-minute time allotment for this study did not provide sufficient time for the actor and participant to build rapport. In the text condition the available time was not even sufficient to relay the story. A longer period of time would be necessary to allow the data to normalize. A longer period is also necessary to make use of IBI data. IBI data analysis runs in 1 minute and 5-minute intervals. There were not enough intervals in the sessions to provide meaningful data.

Artificiality

The actor told the AI character's story to ensure the story was consistent across all conditions. This induced an additional level of artificiality in the human-to-human interactions. The actor was not able to be his genuine self, but rather was sharing someone else's experiences as if they were his own. It was likely that this prevented the actor from connecting at a deeper level with the participants. This could be the reason that the actor and participants did not show any sign of EEG synchrony nor joining up during their dialog.

Real-World Interactions

The study design removed much of the controls that would be experienced in a laboratory-controlled experiment. Rather it functioned as a field study with a wide range of variability since the actor did not have a script. He responded, much like the AI character, to queries from the participants. He could go as deep or as shallow as the participant chose based on their queries. The intent was that the study could lead to greater understanding of the research questions and could guide further research, bridging the gap between lab studies and field experiments to better understand what it takes to engage dialog partners.

Research Questions

What metrics could be applied as an indicator of perceived agency?

1. Biometric Correlates of Social Interactions

Research Question 1: Are Biometric Correlates of Social Interactions appropriate to measure the level of perceived agency based on study condition?

Research Question #1a – Variations in EEG Data

Do variations of brain activity (engagement, wavelength, and workload) in the different study conditions indicate that participants respond to communication mode and/or the interaction partner in differing ways?

EEG Discussion

Cognitive State

The cognitive state of each participant in the face-to-face (condition 1) encounter indicated that two participants experienced distraction and one experienced low engagement. The cognitive state of each participant in the video encounter (condition 2) indicated that one participant experienced high engagement while the other experienced distraction. The cognitive state of each participant interacting with the avatar (condition 3) indicated that all three participants experienced high engagement. The cognitive state of each participant interacting with the AI agent (condition 4) indicated that one participant experienced high engagement and the other experienced low engagement. Finally, the cognitive state of each participant communicating via text (condition 5) indicated that both participants experienced low engagement.

Table 18 summarizes the engagement levels shown in each condition's cognitive state described above. This information was tabulated to provide a sense of which state was most common across conditions.

Table 18 - Summary of Cognitive State Comparison

Face-to-Face (Assigned score)	Video	Avatar	AI Agent	Text
Low Engagement (2) Distraction (1) Distraction (1)	High Engagement (3) Distraction (1)	High Engagement (3) High Engagement (3) High Engagement (3)	High Engagement (3) Low Engagement (2)	High Engagement (3) Low Engagement (2)

Providing each level of engagement a score, with the lowest being distraction (1), the middle being low engagement (2) and the highest being high engagement (3), then averaging the scores of each condition (since there aren't the same number of participant in each condition) provided a subjective measure of the level of engagement experienced in each condition. The results of that calculation are shown in Figure 145. The maximum possible score was 3 with each participant experiencing high engagement. The results indicated that the condition where the cognitive state of the participants was highest was condition 3, avatar. The one with the lowest level of engagement was face-to-face. This calls into question the premise that face-to-face interactions should be considered the "gold standard." Figure 146 shows the same chart with the standard error bar.

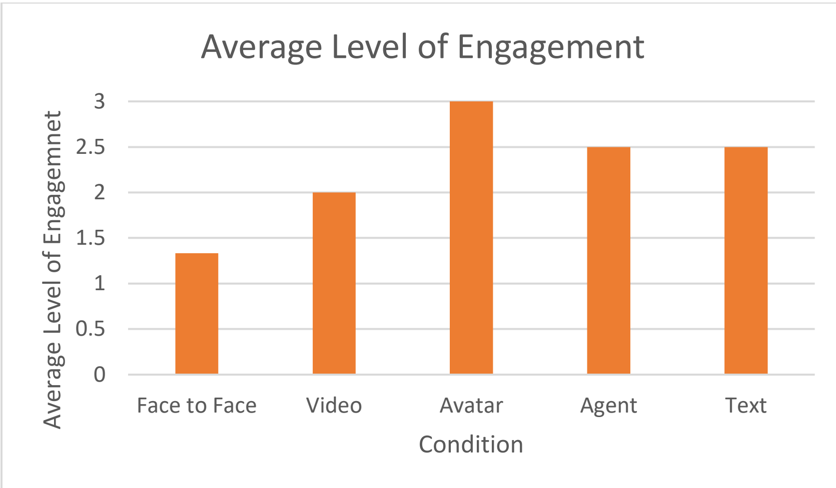

Figure 145 - Average Level of Engagement per Condition

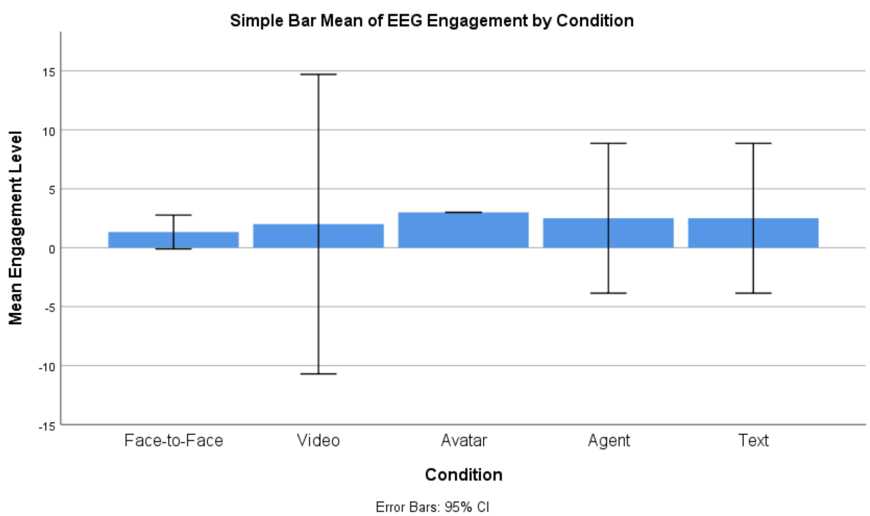

Figure 146 - Average EEG Engagement with Standard Error Bar

This was a biometric measure describing actual activity in the brain. As such, certain questions came to mind. Why were the participants in the face-to-face condition the lowest engaged? This was exactly the opposite of what would be expected. Why were participants in the text condition more engaged than both face-to-face and video? One theory was that people were engaged by the technology rather than the dialog. However, that would not explain why the participants in the text condition were so highly engaged.

Considering engagement as “attention” might shed some light on the results. In the avatar condition (condition 3), the mind might have been fascinated with seeing an avatar move like a human, but at the same time, the mind cannot help but notice slight differences in the avatar’s movement compared to a human’s movement. This might be similar to people experiencing the uncanny valley. A different dynamic might have been at work in the agent condition. Dialog with an agent can be fraught with misinterpretations. It is possible that even at a subconscious level, humans work to find dialog to which the AI can respond. In the open-ended questions of the surveys, one participant stated the (s)he was not certain if the agent was AI, the other participant realized quickly it was AI. Interestingly, the participant who realized it was AI was the one with the highest cognitive state, despite that the participant was falling asleep during the AI’s long monologues. This may have been because his/her brain was activated, even at a subconscious level, in the task of finding ways to communicate with an AI agent. Similarly, the participants in the text condition (condition 5) had their minds activated in the task of interpreting and typing responses, where face-to-face dialog might have been more natural and required less activation.

Wavelengths

The primary two wavelengths, based on amplitude, for each participant are listed in Table 19. Each row showed the top two wavelengths for each participant by condition. The chart showed that each condition had a set of predominant wavelengths. Face-to-face, video, avatar and agent were predominantly theta and alpha. The text condition had delta intertwined with alpha and theta for each participant, suggesting fatigue. Delta also appeared for one participant in the face-to-face condition and one participant in the avatar condition. The conversational elements of this study lent themselves to alpha/theta wavelengths which could be associated with empathy. It would be interesting to see if alpha/theta were predominant wavelengths if the topic of conversation were less emotional. It is unlikely much could be concluded from this chart, but more research might provide insight into the emotional engagement, or distraction of individuals in various conditions.

Table 19 - Each Participant's Predominant Wavelength by Condition

Face-to-Face	Video	Avatar	Agent	Text
Two participants: Alpha/Theta One participant: Alpha/Delta	Both participants: Alpha/Theta	Two participants: Alpha/Theta One participant: Theta/Delta	Both participants Alpha/Theta	Both participants Alpha/Theta/Delta

Workload

It is not clear what role workload played as an indicator of the participant believing their interaction partner had agency. Workload might have indicated how much work a participant applied toward engaging in the dialog. Low workload might have indicated acceptance of a dialog partner as having agency since dialog with another human is natural and automatic.

To better visualize these data, the levels of workload were subjectively broken into levels of High (greater than 60%), Medium (40-60%) and Low (less than 40%). Those values were given numeric representation (high =3, medium=2 and low=1) then they were averaged per condition. This is not a scientific strategy to visualize workload, but it does provide a sense of which condition's participants experienced the least and greatest levels of workload. This information could be used as a strategy to compartmentalize workload in future studies. The results are shown in Figure 147.

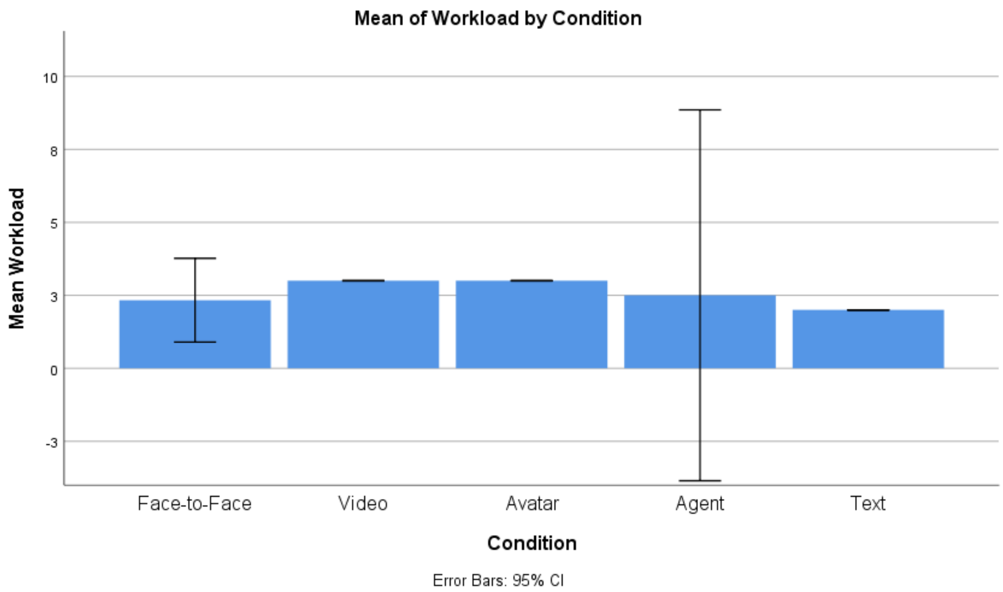

Figure 147 - Workload Comparison across Condition

The data were a bit surprising since the text condition (condition 5) had the lowest workload. This was counter to the premise that people had to work harder to translate their messages to and from text. It was possible that the workload data for the face-to-face condition (condition 1) may have been influenced by two out of three participants having Social Phobia. More research on the relationship between Social Phobia and workload could bring light to this possibility. Further research could provide insight into workload's role in interactions in the future.

Research Question #1b – Variations in Electrodermal Activity

Does arousal, as measured by EDA, indicate that an interaction partner has perceived agency?

EDA Discussion

The EDA data was noisy and unusable, except in two cases. The data that was usable showed a spike in EDA at two times in the study interval. These times correlated with the greatest spikes in the cognitive state data but missed several variations in cognitive state and workload that occurred over the same interval. It was likely that better EDA equipment would have provided different outcomes since other research has shown the value of EDA as a tool to assess engagement with virtual characters, however, that was simply not seen here.

Research Question #1c – Variations in Heart Rate Data

Does the heart's Inter-Beat Interval provide a measure that indicates that an interaction partner has perceived agency?

IBI Discussion

The IBI showed a great amount of individual differences. There was little variability demonstrated as a result of the research conditions. This was shown in Figure 148. Each time measure in the graph was a 30 second epoch. The chart was enlarged to provide greater detail into any patterns. Conditions were in common colors, with condition 1 in orange, 2 in blue, 3 in gray, 4 in green and 5 in yellow. Based on this data, it did not appear that IBI or heart rate variability was a useful measure to indicate if a dialog partner had met the threshold of realism.   It may have been possible that the interval of time when the dialog took place was not sufficient to conduct deeper analysis on the data.

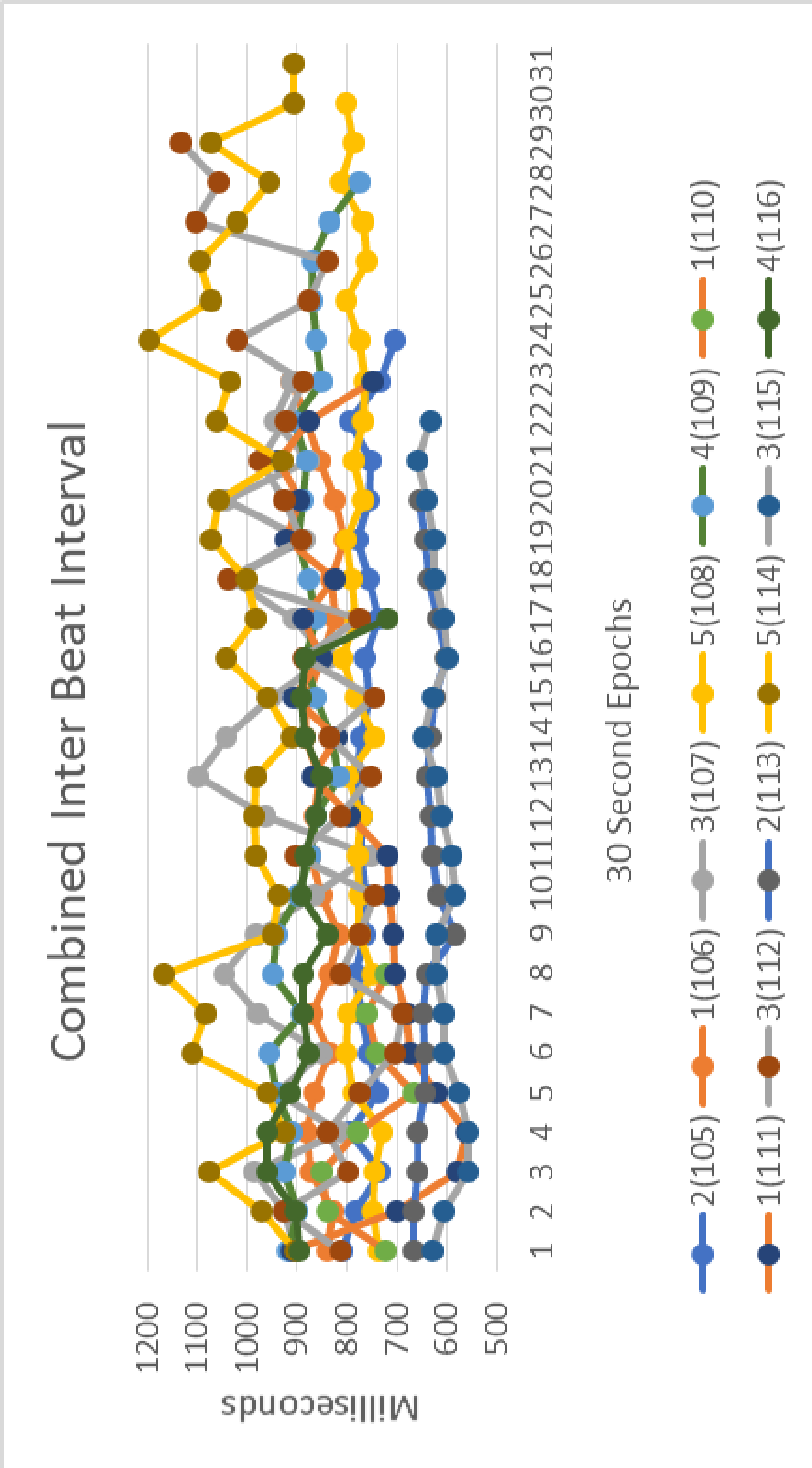

Figure 148 - Combined Inter-Beat Interval

2. *Behavioral Measures*

Research Question 2: Behavioral Measures

Are behavioral measures appropriate to assess perceived agency based on study condition?

Gesture Discussion

There was a clear difference in the amount of gestures per 100 words in the face-to-face condition as compared to the others, as seen in Figure 149. There was one session in the avatar condition that had an even higher relative number of gestures than the face-to-face condition. There seemed to be a great deal of individual differences. For example, the participant in condition 3, who had the greatest number of gestures, was very conscious of how (s)he was seen and how (s)he could modify his/her own gestures to regulate his/her interaction with his/her dialog partner. It was unclear to what extent this affected the actor, since there wasn't any video footage of the actor's gestures during that particular interchange. Other people seemed to use their gestures to encourage the speaker or to indicate that they were along with him on his story-telling journey by nodding. The gestural data between the actor and the participant in Figure 149 have a directly proportional relationship to one another in interchanges where there is data for the actor.

Of course, if an individual knew that they were communicating with a digital character, it would have been unlikely that they would gesture. However, if they believe their dialog partner was real and that the partner could see them, it would be likely that they would augment their dialog with gestures. This seemed to be a rich area for further research.

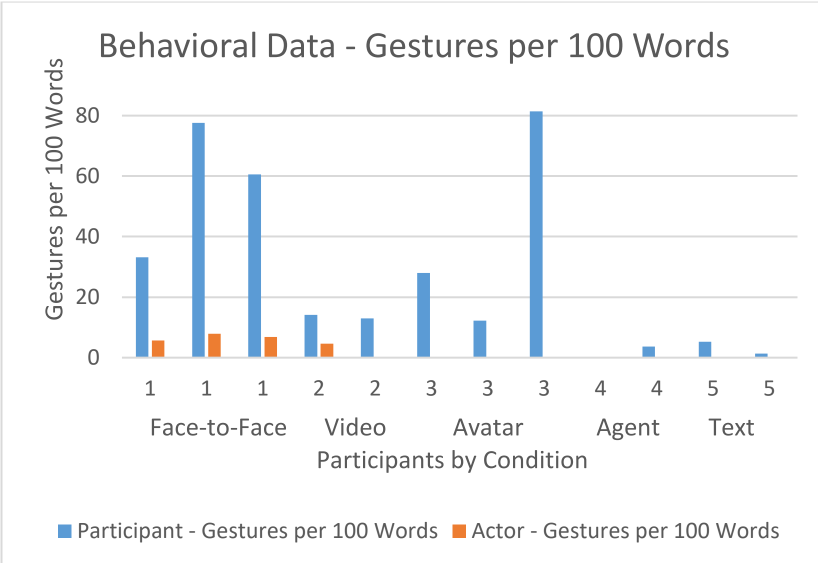

Figure 149 - Gestures per 100 Words

3. *Survey Measures:*

Research Question 3: Does survey data indicate variations in perceived agency based on condition?

Research Question #3a – Presence Questionnaire

Will the Presence Questionnaire (APPENDIX C) provide a measure of perceived agency based on condition?

Presence Discussion

Since each question in this questionnaire explored very different aspects of the sense of presence, examining the results individually provided more insight than the combined score of the questionnaire.

The first question asked "How consistent the dialog experience seemed compared to real-world experiences?" A high score indicated that it was very consistent with the real world. The results, shown in Figure 125 of the FINDINGS section, indicated expected responses based on common experiences in daily lives. Face-to-face interactions are common and expected, as are video meetings. Text is also something most people use on a regular basis. However, both avatar and agent interactions are less common in everyday life and they scored slightly lower. This question did not provide a great amount of insight into participant's experiences. However, adding an open-ended question in future research might provide meaningful insight.

The second question asked "How involved were you in the experience?" Given the results shown in Figure 126 of the FINDINGS section, it appeared that each participant, independent of condition, felt that they were very involved in the experience. However, referring back to the cognitive state data, where some participants were distracted, some experienced low engagement and some experienced high engagement, it was clear that not everyone was as involved as they indicated in this self-report survey. This reinforced that surveys might not provide consistent information in relation to biometric data.

The third question asked "How much delay did you experience between your comments and expected responses?" Responses closer to zero showed no or low lag. Responses closer to 7 indicated significant lag. The results were shown in Figure 127 of the FINDINGS section. Similar to the first question, it was difficult to make sense of the response. It seemed unlikely that participants experienced more lag in the face-to face condition than in the video and avatar conditions. It did make sense that the agent had more lag, however, since the AI had to convert the spoken question to text, then select the appropriate video to play back. The text condition was expected to have greater lag as well since the thoughts needed to be translated to text. This could have been a matter of survey fatigue. This survey was only 4 questions long, but there were three other surveys that were longer and were taken at the same time. It was also possible that the question did not make sense to people in the face-to-face condition, since that was all they experienced. Maybe they considered the set-up time to fit the EEG as part of the lag. Adding an open-ended question to this question in future research might help explain the face-to-face anomaly.

The final question of this questionnaire was "How well could you concentrate on the assigned task?" The responses are shown in Figure 128 of the FINDINGS section. The results indicated that there was little or no difference in the ability to concentrate based on condition. The task, to talk with Jarett about his experience, was consistent across each condition. It appeared that the variations in each condition were not perceived to be an impediment in accomplishing the task.

Research Question #3b – Rapport Questionnaire

Will the Rapport Questionnaire (APPENDIX A) provide a measure of perceived agency based on condition?

Rapport Discussion

The questionnaire results were summative. Figure 129 in the FINDINGS section showed the average scores for each condition which were close to one another. While there was not much difference in the scores on this survey by condition, it may still be a good resource for future studies where the expected participant numbers do reach higher power.

Research Question #3c – Interaction Questionnaire

Will the Interaction Questionnaire (APPENDIX B) provide a measure of perceived agency based on condition?

Interaction Discussion

The original intent was to use the Interaction Questionnaire's summative values, but it was not validated with summative meaning. Rather the questionnaire's greatest value was to dig into each question, along with the open-ended questions to explore what meaning emerged. The questionnaire explored areas of frustration or stress associated with the interchange. For each forced-choice question the participant was asked to provide open-ended responses with the goal of achieving more insight into what the participant was aware of that affected their experience. The open-ended responses were the focus of this discussion.

Across all conditions, responses to the question indicated that the participants were most focused on the content of the dialog with very little focus on the interface strategy. The fact that most participants were focused on the story was positive, indicating that the interface was not foremost in their minds. Those who did express issues brought light to real problems with technology and specific interface strategies. Insights from the open-ended questions of the Interaction Questionnaire for each condition are included in Table 20.

Table 20 - Insights into Open-Ended Responses to Survey

Face-to-Face	Focus was on facial expressions, eye contact, hands shaking, voice modulations
Video	Focus on eyes, not body movement
Avatar	- Difficult to determine attitude and emotions by looking at facial expressions - Didn’t know how I appeared to avatar - Visual and vocal indicators were contrary – laughed but didn’t see that on his face - Arms were straight and not moving
Agent	- Delayed response time/fading “between scenes” - Didn’t answer questions or answered same question - Distracted by body language, fidgeting, shaky voice, hand and leg movements - Couldn’t “break the ice” - Robotic - One participant was falling asleep during interchange

Text	- Time issues with getting to the point - Miscommunication - Slow responses

General Insights

The open-ended questions provided some interesting feedback. For example, one comment about how natural the interaction was pointed out that the most unnatural factor was "probably the study setup. The preparation beforehand skewed my perception of the scenario." The set up for the interchange was laborious. In some cases, it took more than an hour to fit the EEG sensors. The participants provided valuable feedback on the intrusion of the equipment. That might have added a level of artificiality and frustration to each of the interactions.

Avatar Condition Insights

In the avatar condition a human was able to control the avatar's body and facial expression; however, there was still a good amount of information that was lost when transferred to the avatar. While technology allows for more realistic avatars that can be controlled in real-time, the cost for creating and controlling them still prevents widespread availability. High-cost Hollywood movies have made it difficult to differentiate a character that is real or computer generated. However, the level of realism necessary to help an interaction partner read into your expression is simply not available at a reasonable price-point (under $10K). Without being able to model all facial subtleties, some level of communication would be lost, as was pointed out by the participants' responses. This was not indicated in any other data in this study, but provided valuable feedback on where investments might bring about payoffs in the use of this technology.

One participant expressed concern with how he looked to the character controlling the avatar. Had the participant had a sense of how (s)he looked, (s)he might have made adjustments in body language to better aid in the dialog. This was an interesting point that was expressed multiple times by this participant. It was even part of his/her dialog with Jarett, as the participant tried to establish what Jarett could see of him/her. This established that it wasn't just important during a dialog that humans be able to see their dialog partner and have the ability to read subtle expressions to gain understanding, but that the dialog partner also needed to see the human in order to get the added meaning provided by gestures. This participant made great use of his/her eyebrows and other facial expressions during the dialog to express surprise, shock and warmth. This feedback could be very valuable in ensuring future AI has the capacity to pick up these details and respond to them.

Another concern expressed by the participants was that "visual and vocal indicators were often contradictory, such as hearing him laugh but only seeing a faint smile with 'cold eyes' on the avatar." This response emphasized the importance for an avatar to mirror more subtle facial expressions. Detailed expressions can be costly. At the same time, losing a learning partner is costly in different ways. A cost benefit analysis should be used to determine the level of fidelity needed for specific training tasks.

Participants also noted that the avatar's posture was not natural. The actor had to take special precautions to ensure that his movements did not cause the avatar's body to behave in awkward ways so he kept his arms straight. The facial mapping was not perfect and subtle movements like crinkling eyes and moving cheeks were not achieved for this experiment. This is something that would likely be worth the investment to improve on, depending on the tasks to be achieved within the simulated environment.

Agent Condition Insights

While the AI branching and voice recognition used by the ICT DS2A were probably sufficiently advanced if an individual knew they were speaking to AI, these participants were not given that insight. They tried to speak with the AI agent as if he were a person, and this led to moments of frustration. It also did not appear that the branching program tracked if the question had previously been asked and answered causing disconcerting repetition. Additionally, the voice recognition and branching used in the

ICT DS2A were somewhat dated. It is likely that conversational branching has improved since its development. Another interesting factor was that when Jarett responded to detailed questions, his responses were often several minutes long. Since an agent does not have the capacity to assess his listener's continued interest, it might be helpful to break monologues into shorter segments. A human might wait for a nod or "uh-huh" from their listening partner as an indicator of interest and gentle prodding to go on. Without this conversational tool, other strategies might be necessary when interacting with an agent.

One participant was uncomfortable with missing "niceties" that occur in conversation between humans. Basic introductory phrases, or two-way information gathering, might be useful in building rapport. The ICT DS2A was not designed to have a two-way dialog, but rather to respond to questions. Noting the importance of this to a dialog partner helps to define requirements for future AI.

One participant stated that the interaction felt a bit like an animation or robotic. There were moments between videos as the AI interpreted the question asked by the participant where a listening video was shown. Then before a response video was started there was a pause. This sequence might have made the interaction seem animated or robotic, even though the video was of a real person. Both participants felt the pauses made the interaction feel less real.

Research Question #3d – Social Phobia Inventory (SPIN)

The Social Phobia Inventory (APPENDIX D) measured the presence of social anxiety disorder. Will individuals with Social Phobia, as indicated in the SPIN show different results in the above measures as compared to individuals who don't have indications of Social Phobia?

SPIN Discussion

One area where Social Phobia might have had an influence on the results was in workload, since two of the people with social phobia had a higher workload than the other participant in their condition who did not suffer from social phobia. One participant in condition 1 also showed a separation in wavelengths with a jump in delta that might indicate fatigue. This might be related to having to work harder in social situations due to Social Phobia.

The combined workload graph of each participant in condition 1 is shown in Figure 150. Both 110 and 106 experience Social Phobia. It was possible that these two participants, being two-thirds of the face-to-face condition, skewed the data in that condition due to their higher workload. This might have been the case for other data in condition 1 as well.

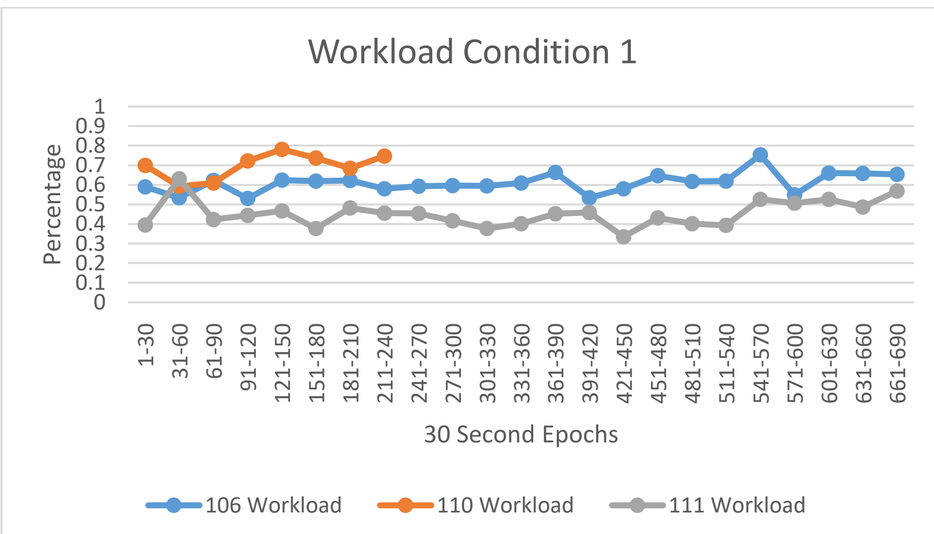

Figure 150 - Workload for Condition 1

The combined workload graph of each participant in condition 4 is shown in Figure 151. Participant 116 experiences Social Phobia, but the workload was not higher than the participant without Social Phobia.

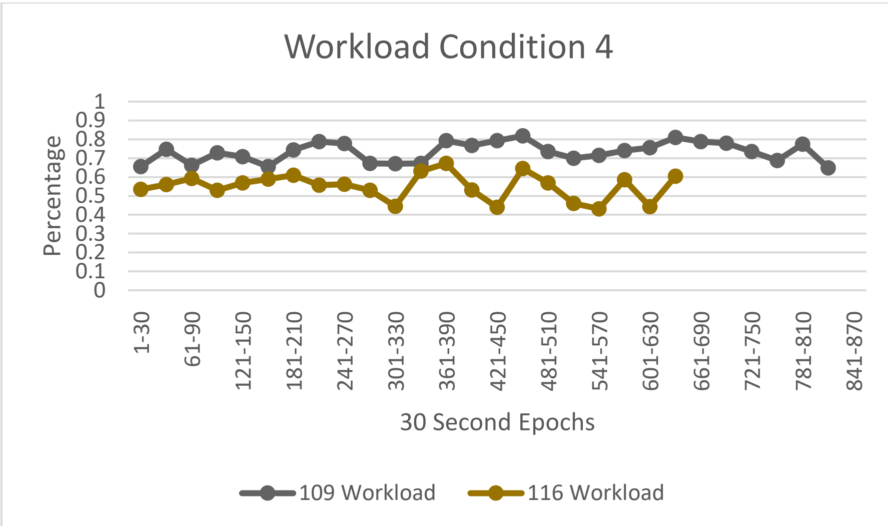

Figure 151 - Workload for Condition 4

Further research should be conducted to determine if people who experience Social Phobia are likely to work harder in dialogs with people than those who don't experience Social Phobia.

Results

The results of this line of research informs requirements for AI of the future in how to decipher when connections are needed and what strategies might aid in making those connections. Ultimately, the outcome will inform future investments in virtual characters to support training tasks. A virtual character that can exceed the threshold necessary to convince an interaction partner that they are real can support various types of training within the Army, across other services as well as in the broader commercial world.

CHAPTER SIX: CONCLUSION

The purpose of this study was focused on understanding what might cause people to engage or disengage with virtual characters. To answer that question, it was important to assess what measurements might indicate that people believe their interaction partner has agency. The results can provide development strategies to improve interactions with virtual characters by informing future requirements.

Despite vast improvements in realism of simulated and game environments, one area still presents a challenge; simulated interpersonal interactions. This study has uniquely contributed to existing research into how individuals react to different communication modalities and interactions with virtual characters that are either fully automated or controlled by a human. This informs future investments in virtual characters by building understanding into what drives humans to engage or disengage. This knowledge facilitates future cost savings and investments in simulated characters and avatars that will support military training tasks and even commercial activities in a manner that will appropriately engage learners and users.

One goal of this research was to explore strategies to reduce the cost of military training by improving the technology used in virtual training environments. Providing realistic training that is available anywhere for a fraction of the price of live training will save money while maintaining military readiness.

An example of a training task that could use realistic virtual characters is advanced situational awareness training. This involves surveillance on a regional area to understand normal patterns of life. Once normal patterns are recognized, a soldier engages with individuals in that area to evaluate how his/her presence disrupts those patterns. During these operations, soldiers are trained to note nonlinguistic body motions, such as eye movements or gestures that might indicate that the individual is untrustworthy or behaving suspiciously. Current training is conducted using live actors over weeks. Virtual training could significantly reduce the cost of conducting this training while maintaining readiness.

Another example of costly live training takes place at the Combat Training Centers (CTCs). Soldiers prepare for months at home station to “deploy” to a CTC and engage in combat against a world-class, unpredictable opposing force. The capabilities described in this research could prepare soldiers for their capstone training by using both AI and human-controlled avatars. Well before deploying, soldiers would receive Situational Reports (SITREPS) that lead to intelligence collection, video surveillance using game-based technology and AI, as well as interviews with real-time human-controlled avatars to build their intelligence portfolios in preparation to arriving at the CTC. Soldiers could conduct key leader engagements, interact with the local population through interpreters, and build advanced situational awareness all through the use of virtual characters well before their live training event occurs.

In order to realize cost savings, virtual characters must model behaviors at the appropriate fidelity to support the training. Multiple theories have explored what factors are important to convince an end-user that a virtual character is real and has agency. Most recently, the Threshold Model of Social Trust/Influence says that certain social factors must be present to exceed a threshold level of realism. Once the threshold is exceeded, individuals will treat that virtual character as if it were real. Factors that influence that perception are agency and fidelity, both behavioral and visual. Based on previous research, the more important of these two factors is agency.

So what measures are used to determine if an individual has gained the sense that a character is real? Past research in a lab setting has used various metrics such as biometric, behavioral, and survey data. This research was not conducted in a lab, but was rather a field study and since it was a novel approach to a real-life situation, a variety of each of these measurement strategies was applied to guide future research plans.

This was a mixed methods study to explore measurement strategies to gain insight into the realism of virtual characters. Neural measures were collected to determine engagement, mood, and workload. Heart rate data were collected to assess stress as a potential indicator of perceived agency. EDA was used to see if activation of the sympathetic nervous system as an indicator of arousal or stress might be an indicator that a virtual character has perceived agency. Behavioral measures, specifically the use of gestures, were applied as a strategy to better understand the research question. Finally, survey data were used to explore rapport, interaction factors, a sense of presence and whether the data could be skewed by individuals with social phobia. Unfortunately, due to limited access to equipment and personnel, fewer participants were put through the study than originally planned.

Future Research

Despite that, the results of this study do provide insight into areas of future research. The EEG results provided various insights into each condition. One example is the comparison of cognitive state which showed that the only condition where each participant demonstrated high engagement was the avatar condition. Two people in the face-to-face condition were in the lowest possible level of engagement with one in the medium range. Each of the other conditions had one person in the highest level of engagement and one in the medium range. Is this simply the result of individual differences? Is there something about the face-to-face interaction that drove people to withdraw? Was there something about the avatar interaction that drew people in? These are all areas of potential future research.

The predominant wavelengths in each condition were alpha and theta. These wavelengths were associated with empathy, emotions and visualization. This made sense since the study scenario involved a dialog about sexual assault. The topic was driven by available agent technology but it had the added benefit of inducing empathy. At the same time, two conditions, face-to-face and avatar each had one participant where delta brainwaves played a significant role. Both participants in the text condition, while having alpha and theta, had a nearly equal amplitude of delta wavelengths. Future research could explore several questions, such as: Does this indicate that they were bored and had sleep onset while waiting for a text message to be received, responded to and returned? Could the other sessions that had such a strong amplitude of delta indicate boredom as well? What drove the participants to get bored?

The workload data showed that the highest workload took place in the avatar condition and the lowest in the text condition. Future research would help to identify the meaning of workload in an interpersonal dialog and if there is a relationship to an individual's sense that their interaction partner is real.

The EDA equipment was discontinued and provided erroneous results. Previous research has made use of EDA as a window into the level of arousal experienced. The expectation was that arousal, and those events that cause it, would provide insight into what events would activate different moods and levels of engagement in the brain activity. Video playback of the dialog would help pinpoint what types of events had the greatest effect on the participant. Future research should create stringent timing synchronization between sensors to explore this concept more fully.

Heart rate data did not produce meaningful information. This was more due to the duration of the dialog than any fault in the equipment. In order to get results from IBI the experimental timeframe needed to be closer to 20 minutes. The sample intervals for the hardware are either 1 minute or 5 minutes. In the 8 to 10-minute sessions during this study, there were not enough samples given the number of participants. Similar research in the future would need to extend the time of the dialog to at least 20 minutes or ensure a larger number of participants, or both.

Behavioral metrics did appear to carry meaning in that the number of gestures per 100 words showed that the face-to-face condition did appear to be the “gold standard.” The number of gestures were higher in the face-to-face condition with some interesting peaks in the avatar condition. The agent and text conditions had a negligible amount of gestural behavior. Despite studies that show that people gesture even while they are on the phone, the gesturing data collected during the dialog is clearly higher when the interaction partner seemed real. This indicated that gestures might be a meaningful measure to indicate a belief that an interaction partner is an actual person. Future research could explore this premise further.

The self-report data proved to contain some very valuable information. There were some meaningful insights that can immediately be applied to virtual characters to improve their sense of realism. Participants pointed out the importance of subtle behavioral cues that they expect from an interactional partner with agency. Shaky voice and hands, eyes that flit about, and moving about in their chairs were some behavioral signs that participants noted.

Additionally, participants noticed that the avatar, which was controlled by a human's movements, did not mirror the actual crinkling of the corner of the eyes as was indicated by the actor's voice. This discontinuity immediately had an effect on the participants. One participant had trouble picking up cues about the interaction partner's attitude toward the participant because of missing cues in the avatar's face.

People who interacted with the AI noted time delays as the program interpreted the spoken question into text and selected the appropriate video footage to play. They also noted the jumps from one animation scene to the next and found it jarring. Insight into what did *not* give the sense of agency helps create an experience that does give the sense of agency.

Future research could help understand what specific stimulus affects an individual’s perception of agency by synchronizing sensors more tightly and possibly using eye-tracking. It might also help to inject events as markers to baseline the cognitive effects of boredom or interest. It would be interesting to assess if a speaker uses more gestures with their dialog partner, would the dialog partner also use more gestures. It would also be interesting to see if that affects engagement. Future research should be applied to understanding the role of workload in an interpersonal dialog. Finally, feedback from the open-ended questions should be implemented to agent and avatar functionality. These changes might influence the outcome of future research.

Closing

The world is getting better at talking to Google, Siri and Alexa, who we know are not real. Future personal assistants may become more useful by becoming more human-like. This has already happened with robo-calls that ask questions and seem to provide human-like responses so that we do not hang up. This same technology could go far in improving training experiences. Who knows, it might even lead us to a training holodeck similar to the one seen on Star Trek. These are some of the thoughts and issues that will emerge as computational power and computer learning improves. Research such as this, along with research that builds upon it, will be the guide to these interchanges in the future.

APPENDIX A - RAPPORT QUESTIONNAIRE

APPENDIX A - RAPPORT QUESTIONNAIRE (von der Putten, Kramer, Gratch, & Kang, 2010)

To be completed by participant - Circle Your Response

Participant Number: ___________ Date: ________________

Session: _________________

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
Jarett created a sense of closeness or camaraderie between us.	1	2	3	4	5
R Jarett created a sense of distance between us.	1	2	3	4	5
In think that Jarett and I understood each other.	1	2	3	4	5
R Jarett communicated coldness rather than warmth.	1	2	3	4	5
Jarett was warm and caring.	1	2	3	4	5
R I wanted to maintain a sense of distance between us.	1	2	3	4	5
I felt I had a connection with Jarett.	1	2	3	4	5
Jarett was respectful to me.	1	2	3	4	5
R I felt I had no connection with Jarett.	1	2	3	4	5
I tried to create a sense of closeness or camaraderie between us.	1	2	3	4	5

R I tried to communicate coldness rather than warmth.	1	2	3	4	5

****R** indicated item is reverse-coded

APPENDIX B - INTERACTION QUESTIONNAIRE

APPENDIX B - INTERACTION QUESTIONNAIRE (Artstein, et al., 2017)

Participant Number: ___________________ Date________________

Session: ____________________

Please place an X in the appropriate box to indicate your response, then provide written responses to the follow-on questions.

1) To what extent were you able to able to effectively communicate with Jarett?

|________|________|________|________|________|________|________|

NOT AT ALL SOMEWHAT COMPLETELY

1.a) Why or Why not?

____

____

____

____

____

2) To what extent were you able to understand Jarett?

|________|________|________|________|________|________|________|

NOT AT ALL SOMEWHAT COMPLETELY

2a) If you couldn't understand him, what was the barrier to understanding?

____

____

____

____

____

3) To what extent were you frustrated during the dialog?

|________|________|________|________|________|________|________|

NOT AT ALL SOMEWHAT COMPLETELY

3a) If you experienced frustration, what was the source?

____

____

____

____

____

4) To what extent did you feel understood?

|________|________|________|________|________|________|________|

NOT AT ALL SOMEWHAT COMPLETELY

4a) What indication did Jarett give that he did or did not understand you?

____

____

____

____

____

5) To what extent did you notice emotional indicators in Jarett?

|________|________|________|________|________|________|________|

NOT AT ALL SOMEWHAT COMPLETELY

5a) What were the emotional indicators you noticed?

____

____

____

____

____

6) To what extent did you trust Jarett?

|________|________|________|________|________|________|________|

NOT AT ALL SOMEWHAT COMPLETELY

6a) Why or why not?

____

____

____

____

____

7) To what extent did you experience stress during the interchange?

|________|________|________|________|________|________|________|

NOT AT ALL SOMEWHAT COMPLETELY

7a) If so, what was the source of stress?

____

____

____

8) To what extent was the interaction natural and believable?

|________|________|________|________|________|________|________|

NOT AT ALL SOMEWHAT COMPLETELY

8a) If the interaction was not natural, what made it feel unnatural?

9) To what extent was Jarett trustworthy?

|________|________|________|________|________|________|________|

NOT AT ALL SOMEWHAT COMPLETELY

9a) What made Jarett seem trustworthy or untrustworthy?

10) To what extent was Jarett persuasive?

|________|________|________|________|________|________|________|

NOT AT ALL SOMEWHAT COMPLETELY

10a) What made Jarett seem persuasive or unpersuasive?

____

____

____

____

____

11) To what extent was Jarett's voice natural, if you heard him?

|________|________|________|________|________|________|________|

NOT AT ALL SOMEWHAT COMPLETELY

11a) What made the voice seem natural or unnatural?

____

____

____

____

____

12) To what extent did Jarett's movement seem natural, if you could see him?

|________|________|________|________|________|________|________|

NOT AT ALL SOMEWHAT COMPLETELY

12a) What made the movement seem natural or unnatural?

____

____

____

____

____

13) To what extent did you like Jarett?

|________|________|________|________|________|________|________|

NOT AT ALL SOMEWHAT COMPLETELY

14) To what extent did you dislike the Jarett?

|________|________|________|________|________|________|________|

NOT AT ALL SOMEWHAT COMPLETELY

13/14a) What did you like or dislike about Jarett?

____

____

____

____

____

15) To what extent did you enjoy the interchange?

|________|________|________|________|________|________|________|

NOT AT ALL SOMEWHAT COMPLETELY

15a) What did you find enjoyable or unenjoyable about the interchange?

APPENDIX C - PRESENCE QUESTIONNAIRE

APPENDIX C - PRESENCE QUESTIONNAIRE (Witmer & Singer, 1998)

Participant: ________________ Date: ______________

Session: ___________________

Characterize your experience, by marking an "X" in the appropriate box of the 7-point scale, in accordance with the question content and descriptive labels. Please consider the entire scale when making your responses, as the intermediate levels may apply. Answer the questions independently in the order that they appear. Do not skip questions or return to a previous question to change your answer.

1. How much did your experience seem consistent with your real world experiences?

|________|________|________|________|________|________|________|

NOT	MODERATELY CONSISTENT	VERY CONSISTENT

2. How involved were you in the experience?

|________|________|________|________|________|________|________|

NOT INVOLVED	MILDLY INVOLVED	COMPLETELY ENGROSSED

3. How much delay did you experience between your comments and expected responses?

|________|________|________|________|________|________|________|

NO DELAYS	MODERATE DELAYS	LONG DELAYS

4. How well could you concentrate on the assigned task?

|_________|_________|_________|_________|_________|_________|_________|

NOT AT ALL SOMEWHAT COMPLETELY

APPENDIX D - SOCIAL PHOBIA INVENTORY (SPIN)

APPENDIX D - SOCIAL PHOBIA INVENTORY (SPIN) (Connor, et al., 2000)

Usage Agreement does not allow the SPIN to appear in a form that allows it to be accessible to the public.

For access please contact mail@cd-risc.com

APPENDIX E - DEMOGRAPHICS QUESTIONNAIRE

APPENDIX E - DEMOGRAPHICS QUESTIONNAIRE

Participant Number: __________ Date: _______________

Session: ______________

1. **Age:** __________
2. **Gender** Circle or fill in response
   -Female
   -Male
   -Other __________
   -Prefer not to answer
3. **Ethnicity:** Circle or fill in response
   -White
   -Hispanic or Latino
   -Black or African American
   -Native American or American Indian
   -Asian/Pacific Islander
   Other ______________________
   -Prefer not to answer
4. **Education:** Circle One
   - Minimal or no school
   -Some high school, no diploma
   -High school graduate, diploma or the equivalent (i.e. GED)
   -Some college credit, no degree
   -Trade/technical/vocational training
   -Associate degree
   -Bachelor's degree
   -Master's degree
   -Professional degree
   -Doctorate degree
   -Prefer not to answer
5. **Marital Status:** Circle One
   -Single, never married
   -Married or domestic partnership
   -Widowed
   -Divorced
   -Separated
   -Prefer not to answer

6. **Professional or Employment Status:** Circle one
   -Employed full time
   -Employed part time
   -Out of work
   -Homemaker
   -Student
   -Military
   -Retired
   -Unable to work
   -Prefer not to answer
7. **3-D Immersive Game Usage, includes games that make use of graphic representations of 3D space:** Circle one
   - Never play these types of games
   - Have played, but rarely (once or twice a year)
   - Play once or twice a month
   - Play once or twice a week
   - Play frequently throughout the week
   - Play daily
   - Prefer not to answer
8. **Have you experience sexual assault or sexual abuse**
   - Yes
   - No
9. **Are you studying to be or currently occupied in a mental health profession**
   - Yes
   - No

APPENDIX F - ELECTROENCEPHALOGRAPHY QUESTIONNAIRE

Appendix F - Electroencephalogram (EEG) Factors Questionnaire

To be completed by participant

Participant Number: __________ Date: _______________

Session: ________________

1. How recently have you consumed caffeine? __________________(in hours)
    a. What did you consume, and how much?

____

2. How recently have you used a nicotine product? __________________(in hours)
    a. What did you consume, and how much?

____

3. How recently have you consumed alcohol? __________________(in hours)
    a. What did you consume, and how much?

____

4. How recently have you consumed melatonin? __________________(in hours)
    a. How much did you consume?

____

5. How recently have you consumed Marijuana or a component of Marijuana? _________________________ (in hours)
    a. What did you consume, and how much?

____

6. How recently have you consumed other drugs? _________________________ (in hours)
    a. What did you consume, and how much?

____

7. How much sleep did you have last night? __________________(in hours)

8. Do you feel that you are well-rested now? (Yes/No) (circle one)

9. How recently did you eat? __________________(in hours)

    a. Would you consider the meal: light, medium or heavy? (circle one)

10. How would you rate your current stress level? (high/medium/low) (circle one)

11. Do you currently take dietary supplements? If so, what do you take

    ____

    How much do you take? ___________________________

    How often? ___________________________

12. Which hand is your dominant hand? (Circle One)

    Right Left Ambidextrous

APPENDIX G - INFORMED CONSENT

APPENDIX G – INFORMED CONSENT

Title of research study: Interactions Between Humans, Virtual Agent Characters and Virtual Avatars

Informed Consent

Principal Investigator(s):

Tamara Griffith

Co-Investigators or Sub-Investigator(s):

Cali Fidopiastis, PhD

Faculty Supervisor:

Patricia Bockelman-Morrow, PhD

Sponsor: U.S. Army Research Lab

Investigational Site(s): University of Central Florida, Partnership II

Why am I being invited to take part in a research study?

We invite you to take part in a research study because you are over 18 years of age and speak English as your primary language and you have no history of sexual assault or sexual abuse.

What should I know about a research study?

Someone will explain this research study to you.

Whether or not you take part is up to you.

You can choose not to take part.

You can agree to take part and later change your mind.

Your decision will not be held against you.

You can ask all the questions you want before you decide.

This study involves graphic language about sexual assault/sexual abuse.

Who can I talk to?

If you have questions, concerns, or complaints, or think the research has hurt you, talk to the research team at tami.griffith@knights.ucf.edu or pbockelm@ist.ucf.edu.

This research has been reviewed and approved by an Institutional Review Board ("IRB"). You may talk to them at 407-823-2901or irb@ucf.edu if:

Your questions, concerns, or complaints are not being answered by the research team.

You cannot reach the research team.

You want to talk to someone besides the research team.

You have questions about your rights as a research subject.

You want to get information or provide input about this research.

Why is this research being done?

This study seeks to better understand interpersonal interaction in a modeling and simulation environment through the examination of cognitive markers during an interactions between two people. You will experience one of the following interactions:

Interacting with an individual while sitting in front of one another,

Interacting via video teleconference,

Interacting through a human-controlled avatar,

Interacting with a computer-controlled character

Or via text messaging.

This study seeks to examine interpersonal interactions in a modeling and simulation environment through the examination of cognitive markers during interactions between two people. The outcome will inform future investments in virtual characters to support training tasks.

The research is intended to evoke an emotional response and empathy/sympathy with the character during the interaction. The Defense Equal Opportunity Management Institute (DEOMI) is exploring this technology to train human resource personnel to recognize indications of sexual harassment, Post-Traumatic Stress Disorder (PTSD) and depression. The U.S. Army is interested in modeling realistic sympathetic characters in military simulations. The long-term goal is that the characters are sufficiently realistic to portray characteristics such as dishonesty, fear, friendliness and distrust.

How long will the research last?

We expect that you will be in this research study for 30 minutes.

How many people will be studied?

We expect about 173 people will be recruited to participate in this research study.

What happens if I say yes, I want to be in this research?

You will be asked to complete a demographic questionnaire.

Your forehead and fingers will be cleaned, then you will have an Electroencephalography (EEG) device placed upon your head by the research team to collect electrical activity along your scalp. You will need to complete a baselining task after the device has been emplaced (Figure 1). You will also have a SCR system connected to your hand (Figure 2). Neither of these devices are likely to cause discomfort.

Figure 1 - B-Alert X10 and X24 EEG System

Figure 2 - Skin Conductance Response sensor

Your experiences will be video recorded during this study. If you do not want to be recorded, you will not be able to be in the study. The recording will be kept in a locked, safe place. The recording will be erased or destroyed two years after the completion of the study. One video recording device will be facing you and one will face what you are viewing.

You will experience one of five different interactive tasks after which you will be asked to complete surveys about the experience. The experiences you have will be randomly selected and you will only experience one.

The proctor will introduce you to your interaction partner and will leave the room to the next room.

You will interact with one individual who will respond to your questions about his experience of sexual abuse or assault. Your task is to gain an understanding of the event, inviting the person to share what happened and their feelings about it. You will be answering questions about the event and the person's emotional state after the meeting is over. The interaction will continue for between 10 and 12 minutes. The proctor will step in to end the session.

This study involves graphic language about sexual assault/sexual abuse.

After the interaction you will be asked to complete a Rapport Survey, a Communication and Interaction Survey and a Presence Survey. The surveys will take roughly 10 minutes.

Following the completion the surveys, the EEG and SCR will be removed.

When the final survey is completed and the EEG and SCR are removed, you will have a discussion with the proctor to process what you experienced during the sessions.

If the experience of discussing sexual assault or abuse has troubled you, you will have the opportunity to share this with the proctor and receive a brochure with contact information for additional assistance from UCF Counseling and Psychological Services (CAPS) facility or the RESTORES laboratory in the Psychology Department of the University of Central Florida.

After the debriefing you will be finished with your participation.

What happens if I do not want to be in this research?

Participation in research is completely voluntary. You can decide to participate or not to participate. You are free to withdraw your consent and discontinue participation in this study at any time without prejudice or penalty. Your decision to participate or not participate in this study will in no way affect your enrollment, grades, employment or your relationship with the individuals who may have an interest in this study.

What happens if I say yes, but I change my mind later?

You can leave the research at any time it will not be held against you. Your data will be removed from the study results and the video recordings will be deleted.

Is there any way being in this study could be bad for me?

The dialog topic selected for this research was driven by the state-of-the-art artificial intelligence functionality available at the time of writing. Specifically, the DS2A application developed by the University of Southern California (USC) Institute for Creative Technologies (ICT). DS2A is an application that introduces participants to Specialist Jarett Wright. He is an actual survivor of sexual assault while in the Army. Jarett is a photorealistic video representation and responds to various questions and displays a range of animations. This means that you will be exposed to 10-12 minutes of dialog about sexual assault. The proctor has received rudimentary training to help you unpack the experience and discuss it. This is done through a discussion at the end of the study, where you are asked details from the interviews. If there are indications that you are suffering from a reaction to the content, the proctor will provide contact information to the UCF Counseling and Psychological Services (CAPS) facility and the RESTORES Laboratory.

There is no more than usual risk when wearing the EEG headset. There may be some discomfort in wearing the EEG. The risk of allergy is due to the gel, which is similar in content to human sweat. The risk estimate for allergic reaction to the EEG gel is less than 1% and all risks including discomfort are reversible. Overall risk estimate: Frequency (,1%), Severity (Minimal to no severity), Reversibility (>99%).

The psychophysiological sensors used in this study pose no more than minimal risk to the participant. The devices are disinfected and cleaned after each use. As commercially available devices, there has been no available documentation of high-risk situations. If any adverse situations should occur, for example, skin irritation, the incident will be evaluated to ensure that this instance is isolated and that there is no damage to the equipment that may be the cause of such an issue. If the equipment is found to be faulty, then a backup system will be used.

What happens to the information collected for the research?

Efforts will be made to limit the use and disclosure of your personal information, including research study and any records, to people who have a need to review this information. We cannot promise complete secrecy. Organizations that may inspect and copy your information include the IRB and other representatives of this organization.

These data will be password protected and only accessible by CITI-trained researchers who have been approved by the UCF IRB to have access to the study documents. Identifiable data, such as video recordings, will be stored for 2 years for reference. De-identified data will be stored for a minimum of 5 years. Participant signatures will be stored for 5 years in a separate location from the data.

What else do I need to know?

This researcher conducting this research is an employee of the United States Army Research Laboratory. The technology and resources being utilized in this research study are provided by the United States Army Research Laboratory and the results are intended to support the development of realistic interactive virtual characters and avatars for training in the future.

Signature Block for Capable Adult

Your signature documents your permission to take part in this research.

____ ____________________

Signature of subject   Date

Printed name of subject

Signature of person obtaining consent

Date

Printed name of person obtaining consent

APPENDIX H - DEBRIEFING STATEMENT

APPENDIX H - DEBRIEFING STATEMENT

For the study entitled:

"Interactions Between Humans, Virtual Agent Characters and Virtual Avatars"

Dear Participant;

During this study, you spoke with a survivor of sexual assault. The information shared with you was based on the experiences of Specialist Jarett White as recorded by the University of Southern California (USC) Institute for Creative Technology (ICT). If you spoke to an individual different than the one shown below, you spoke with an actor portraying Jarett's experience as if they were his own. We did not tell you this before you participated in the study because knowing you were interacting with an actor rather than the true individual who experienced this event may have affected the level of engagement you would have with that person.

You are reminded that your original consent document included the following information: Participation in research is completely voluntary. You can decide to participate or not to participate. You are free to withdraw your consent and discontinue participation in this study at any time without prejudice or penalty. Your decision to participate or not participate in this study will in no way affect your enrollment, grades, employment or your relationship with the individuals who may have an interest in this study.

If you have any concerns about your participation or the data you provided in light of this disclosure, please discuss this with us. We will be happy to provide any information we can to help answer questions you have about this study.

Now that you know the true nature of the study, you have the option of having your data removed from the study. Please contact the PI if you do not want your data to be used in this research and it will be withdrawn.

Study contact for questions about the study or to report a problem: If you have questions, concerns, or complaints or think the research has hurt you please contact: Tami Griffith, Graduate Student, Modeling and Simulation Department at 407-384-3636 or via email at tami.griffith@knights.ucf.edu or Dr. Patricia Bockelman-Morrow, Faculty Supervisor, Modeling and Simulation Department at (407) 822-2115 or by email at pbockelm@ist.ucf.edu.

IRB contact about your rights in the study or to report a complaint: Research at the University of Central Florida involving human participants is carried out under the oversight of the Institutional Review Board (UCF IRB). This research has been reviewed and approved by the IRB. For information about the rights of people who take part in research, please contact: Institutional Review Board, University of Central Florida, Office of Research & Commercialization, 12201 Research Parkway, Suite 501, Orlando, FL 32826-3246 or by telephone at (407) 823-2901.

If you have experienced distress as a result of your participation in this study, referral to the University of Central Florida Counseling and Psychological Services (CAPS) facility or to the RESTORES laboratory for free support may be available if needed.

Please again accept our appreciation for your participation in this study.

APPENDIX I - APPROVAL OF HUMAN RESEARCH

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Tamara Griffith** and Co-PIs: **Dr. Cali Michael Fidopiastis, PhD, Patricia B Morrow**

Date: **December 14, 2018**

Dear Researcher:

On 12/12/2018 the IRB approved the following human participant research until 12/11/2019 inclusive:

Type of Review:	UCF Initial Review Submission Form Full Board Review
Project Title:	Interactions between humans, virtual agent characters, and virtual avatars
Investigator:	Tamara Griffith
IRB Number:	SBE-18-14351
Funding Agency:	Army Research Laboratory(ARL)
Grant Title:	N/A
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30 days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 12/11/2019, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a signed and dated copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

This letter is signed by:

IRB Chair

APPENDIX J - DEPARTMENT OF DEFENSE INSTITUTIONAL AGREEMENT FOR IRB REVIEW

Department of Defense
Human Research Protection Program

DOD INSTITUTIONAL AGREEMENT
FOR INSTITUTIONAL REVIEW BOARD (IRB) REVIEW

BETWEEN

The University of Central Florida,
By and on behalf of its Board of Trustees

AND

U.S. Army Combat Capabilities Development Command Soldier Center (CCDC – Soldier Center)

AND

U.S Army Combat Capabilities Development Command Army Research Laboratory (CCDC – ARL)

PART 1
INSTITUTION INFORMATION

This DOD Institutional Agreement for IRB Review describes the responsibilities of the engaged institution and the institution with the IRB. This Agreement, when signed, becomes part of the engaged institution's Federal Assurance for the Protection of Human Research Subjects approved by DOD (and may become part of the Federalwide Assurance (FWA) approved by the Department of Health and Human Services (DHHS)).

A. Engaged Institution Relying on the IRB

Name: Combat Capabilities Development Command Soldier Center (CCDC – Soldier Center)
DOD Assurance Number: DOD A20124
DHHS FWA Number (if applicable): N/A
DOD Addendum to the DHHS FWA Number (if applicable): N/A

B. Institutions Supplying the IRB Services

Name: University of Central Florida (UCF)
DOD Assurance Number (if applicable): N/A
DOD IRB Number (if applicable): N/A
DHHS FWA Number: 00000351

IAIR-UCF-ARL-NSRDEC_2019-02-11_1

DHHS IRB Number: 00001138
DOD Addendum to the DHHS FWA Number (if applicable): N/A

Name: Combat Capabilities Development Command Army Research Laboratory (CCDC – ARL)
DOD Assurance Number: DOD A10038
DOD IRB Number (if applicable): N/A
DHHS FWA Number: N/A
DHHS IRB Number: N/A
DOD Addendum to the DHHS FWA Number (if applicable): N/A

C. Scope

This Agreement applies to the following research conducted by the engaged institution:

[ ] A single research protocol only (list title and other identifying information):

[X] A group of research protocols (describe here or attach list):

For all studies supported by CCDC – Soldier Center where both UCF and CCDC – Soldier Center employees are engaged:

1) The UCF IRB will serve as the IRB of record for all studies where the principal investigator is engaged in the research and is an employee of UCF.
2) The UCF IRB will serve as the IRB of record for all studies where the principal investigator engaged in the research is a CCDC – Soldier Center employee and the research being conducted is in support of the CCDC – Soldier Center employee's academic degree from UCF, e.g. a Ph.D. dissertation.
3) The CCDC – ARL IRB will serve as the IRB of record for all studies where the principal investigator engaged in the research is a CCDC – Soldier Center employee and the study is not being conducted in support of the CCDC – Soldier Center employee's academic degree from UCF.

This Agreement does not eliminate the need for the Human Research Protection Official Review.

[ ] All research performed by this institution.

D. Effective Dates

This Agreement is effective as of the date approved and signed by the DOD Component Designated Official and expires on the date listed in the DOD approval document.

PART 2
INSTITUTIONAL RESPONSIBILITIES

All institutions are responsible for ensuring that their personnel (i.e., the Institutional Official, the IRB, IRB office staff, investigators and research staff, and any other personnel supporting research covered under this Agreement) act in accordance with all applicable federal, state and local laws and regulations (e.g., Title 32 Code of Federal Regulations Part 219 (32 CFR 219); Title 10 United States Code Section 980 (10 USC 980); DOD Directives and Instructions (e.g., DODD 3216.02); 45 CFR Part 46 (Subparts B, C, and D as made applicable by DODD 3216.02); DOD Component policies; and the Food and Drug Administration regulations and guidance (e.g., 21 CFR Parts 50, 56, 312, and 812) where applicable in addition to the terms and conditions of the organizations' DOD Assurance and/or their DHHS FWA.

Specific DOD Component requirements are stated in Part 3 of this document.

All institutions will permit, upon request, the inspection of any facilities used in support of the activities described in the "Scope" and other research areas by federal agencies responsible for oversight of human research protection and proper management of the research within the scope of this agreement.

A. The Institutional Official of the Engaged Institution Relying on the IRB will:

1. Ensure that all institutional personnel involved in the research (covered within the scope of this agreement) have completed education and training requirements.

2. Verify that scientific review of the research protocol has been conducted and that the IRB considered the feedback from the scientific review.

3. Verify that the IRB has reviewed the research protocol in accordance with DOD requirements, including those identified in the research contract or agreement.

4. Ensure institutional personnel comply with requirements and oversight established by the IRB.

5. Ensure institutional personnel follow the approved research protocol.

6. Ensure institutional personnel report to the IRB and DOD: (a) unanticipated problems involving risks to subjects or others; (b) serious or continuing non-compliance; (c) suspension or termination of IRB approval; and (d) any other events or circumstances requiring notification.

7. Ensure institutional personnel maintain current copies of the IRB approved research protocol (initial review, continuing review, amendments, adverse event reports, and final report), all communications with the IRB, this Agreement, and other relevant information in accordance with DOD record keeping requirements.

8. Verify the IRB has the expertise and policies and procedures needed to review and oversee the research submitted by the institution (in accordance with 32 CFR 219.107, §.103(b)(3), and §.115).

B. The Institution Supplying the Reviewing IRB will:

1. Verify that personnel involved in the research have completed required education and training for the protection of human research subjects.

2. Verify that the IRB is properly constituted for reviewing the study.

3. Fulfill the IRB responsibilities identified in the engaged institution's assurance.

4. Provide the Institutional Official of the engaged institution with information about the IRB, such as a list of IRB members or expertise and the written procedures for executing IRB responsibilities in accordance with paragraph A.8 above.

5. Provide to the engaged institution conducting the research and the Principal Investigator(s) a copy of the IRB review and determinations concerning the research (e.g., IRB minutes or other appropriate documents).

6. Maintain current copies of the IRB approved research protocol (initial review, continuing review, amendments, adverse events reports, and final report), all communications with the institution, this Agreement, and other relevant information in accordance with DOD Component record-keeping requirements.

C. Amendments and Termination

1. This Agreement may be modified, cancelled, or renegotiated upon mutual consent, at any time through an amendment signed by authorized representatives of the organizations. A decision to amend or terminate will be submitted to the DOD Component Designated Oversight Official.

2. The DOD Component Designated Official is not obligated to approve this Agreement.

PART 3
DOD COMPONENT REQUIREMENTS

A. This institution will comply with the requirements of the DOD Component issuing this Agreement. These requirements are identified in Part 3, paragraph B. DOD Components may require that other research, not specifically identified by 32 CFR 219, also comply with the terms of this Agreement (32 CFR 219.101(d)).

B. When this institution conducts research supported by or in collaboration with an organization of another DOD Component, this institution must comply with the policies and procedures of that organization. The requirements of selected DOD Components are identified below:

Department of the Army

- AR 70-25 Use of Volunteers as Subjects of Research, 25 January 1990
- AR 40-38, Clinical Investigation Program, 1 September 1989
- AR 40-7, Use of Investigational Drugs in Humans and the Use of Schedule I Controlled Drug Substances, 4 January 1991

Department of the Navy

- SECNAVINST 3900.39D of 6 November 2006

Department of the Air Force

- Department of Defense Instruction 3216.02_Air Force Instruction 40-402, Protection of Human Subjects and Adherence to Ethical Standards in Air Force Supported Research, 10 September 2014

Office of the Under Secretary of Defense (Personnel and Readiness)

- Research Regulatory Oversight Office Human Research Protection Program Operating Instruction, 29 September 2014

National Geospatial-Intelligence Agency

- Policy Notice 3216.1

National Security Agency

- NSA/CSS Policy 10-10, "Protecting Human Subjects of Research", dated 19 April 2016

Defense Threat Reduction Agency (DTRA)

- DTRA Directive 3216.01, dated 9 June 2010, updated 18 March 2015
- DTRA Instruction 3216.02, dated 21 October 2011, updated 31 July 2015

Defense Advanced Research Projects Agency (DARPA)

- DARPA Instruction No. 66, "Protection of Human Subjects in Research," dated 19 November 2013

PART 4
INSTITUTIONAL AGREEMENT

A. Engaged Institution Relying on the External IRB

1. Institutional Signatory Official at the Engaged Institution (CCDC – Soldier Center)

Acting in an authorized capacity on behalf of this institution and with an understanding of the institution's responsibilities under its assurance, I assure protections for human subjects as specified above.

Signature: Date: 8 MARCH 19

Name: Douglas A. Tamilio
Rank/Grade: SES
Institutional Title: Director, Combat Capabilities Development Command Soldier Center
Telephone Number: 508-233-4700
FAX Number: N/A
Email Address: douglas.a.tamilio.civ@mail.mil
Mailing Address: U.S. Army Combat Capabilities Development Command Soldier Center
ATTN: RDNS-D
General Greene Ave.
Natick, MA 01760

2. Primary Contact for Human Research Protection at the Engaged Institution (CCDC – Soldier Center)

Name: Kenneth J. Desabrais, Ph.D.
Rank/Grade: GS-14
Institutional Title: Human Protections Administrator
Telephone Number: 508-233-5946
FAX Number: N/A
Email Address: kenneth.j.desabrais.civ@mail.mil
Mailing Address: U.S. Army Combat Capabilities Development Command Soldier Center
ATTN: RDNS-CS
General Greene Ave.
Natick, MA 01760

B. Institutions with the Reviewing IRB

1. Reviewing IRB Chair Agreement (UCF)

Acting in an authorized capacity on behalf of the IRB and with an understanding of the institution's responsibilities under this assurance, I assure protections for human subjects as specified above.

Signature: Kiminobu Sugaya Date: 2/15/2019

Name: Kiminobu Sugaya, Ph.D.
Rank/Grade: N/A
Institutional Title: IRB Chair, Professor, UCF Biomolecular Science Center
Telephone Number: (407) 266-7045
FAX Number: (407) 823-3299
Email Address: ksugaya@ucf.edu
Mailing Address: 12201 Research Parkway, Suite 501
Orlando, FL 32826-3246

2. Institutional Official of Institution with the Reviewing IRB (UCF)

I am aware that my IRB is entering into this agreement.

Signature: Date: February 15, 2019
Debra R. Reinhart

Name: Debra Reinhart
Rank/Grade: N/A
Institutional Title: Pegasus Professor and Associate VP for Research and Scholarship
Telephone Number: (407) 823-2315
FAX Number: (407) 882-1156
Email Address: debra.reinhart@ucf.edu
Mailing Address: 4365 Andromeda Loop N.
Orlando, FL 32816

3. Primary Contact for Human Research Protection at the Institution with the Reviewing IRB (UCF)

Name: Gillian Morien, BA, CIP
Rank/Grade: N/A
Institutional Title: IRB Manager
Telephone Number: (407) 823-2508
FAX Number: (407) 823-3299
Email Address: Gillian.Morien@ucf.edu
Mailing Address: 12201 Research Parkway, Suite 501
Orlando, FL 32826-3246

4. Reviewing IRB Chair Agreement (CCDC – ARL)

Acting in an authorized capacity on behalf of the IRB and with an understanding of the institution's responsibilities under this assurance, I assure protections for human subjects as specified above.

Signature: Diane M Ungvarsky Date: 12 Feb 2019

Name: Diane Ungvarsky
Rank/Grade: N/A
Institutional Title: Army Research Laboratory IRB Co-Chair
Telephone Number: (913) 684-7663
FAX Number: N/A
Email Address: diane.m.ungvarsky.civ@mail.mil
Mailing Address: U.S. Army Research Laboratory
Human Research & Engineering Directorate
ATTN: CCRL-HR
7101 Mulberry Point, Building 459
Aberdeen Proving Ground, MD 21005-5425

5. Institutional Official of Institution with the Reviewing IRB (CCDC – ARL)

I am aware that my IRB is entering into this agreement.

Signature: Date: 12 Feb 2019

Name: Yun-Sheng "Jessie" C. Chen
Rank/Grade: ST
Institutional Title: Deputy Institutional Official Army Research Laboratory
Telephone Number: (410) 278-1447
FAX Number: (410) 278-5997
Email Address: yun-sheng.c.chen.civ@mail.mil
Mailing Address: U.S. Army Research Laboratory
Human Research & Engineering Directorate
ATTN: CCRL-HR
Aberdeen Proving Ground, MD 21005-5425

6. Primary Contact for Human Research Protection at the Institution with the Reviewing IRB (CCDC – ARL)

Name: Theresa M. Straut
Rank/Grade: N/A
Institutional Title: Human Protection Administrator
Telephone Number: (410) 278-5928
FAX Number: N/A
Email Address: theresa.m.straut.civ@mail.mil
Mailing Address: U.S. Army Research Laboratory
Human Research & Engineering Directorate
ATTN: CCRL-HR
7101 Mulberry Point, Building 459
Aberdeen Proving Ground, MD 21005-5425

APPENDIX K - UNIVERSITY OF CENTRAL FLORIDA IRB APPROVAL

Institutional Review Board
FWA00000351
IRB00001138Office of Research
12201 Research Parkway
Orlando, FL 32826-3246

UNIVERSITY OF CENTRAL FLORIDA

APPROVAL

March 28, 2019

Dear Tamara Griffith:

On 3/28/2019, the IRB reviewed the following submission:

Type of Review:	Modification
Title:	Interactions between humans, virtual agent characters, and virtual avatars
Investigator:	Tamara Griffith
IRB ID:	MOD00000154, SBE-18-14351
Funding:	None
Grant ID:	None
IND, IDE, or HDE:	None
Documents Reviewed:	• Consent, Category: Consent Form; • Recruitment Advertisement, Category: Recruitment Materials; • Presence Questionnaire, Category: Survey / Questionnaire; • Demographic Quesionnaire, Category: Survey / Questionnaire; • Social Phobia Questionnaire, Category: Survey / Questionnaire; • Faculty Advisor Review, Category: Faculty Research Approval; • Rapport Questionnaire, Category: Survey / Questionnaire; • Interaction Questionnaire, Category: Survey / Questionnaire; • Debriefing Statement, Category: Debriefing Form; • Protocol, Category: IRB Protocol; • Application Form, Category: Other;

The IRB approved the protocol from 3/28/2019 to 12/11/2019. This study was deemed by the Full Board as Greater than Minimal Risk.

In conducting this protocol, you are required to follow the requirements listed in the Investigator Manual (HRP-103), which can be found by navigating to the IRB Library within the IRB system.

The PI is responsible to have their research monitor, Patricia Bockelman, supervise the research. She will witness at least one participant and one consent process. She will also oversee the management of research data and consult with the Ms. Griffith throughout the process.

If you have any questions, please contact the UCF IRB at 407-823-2901 or irb@ucf.edu. Please include your project title and IRB number in all correspondence with this office.

Sincerely,

Gillian Morien
Designated Reviewer

APPENDIX L - HUMAN RESEARCH PROTECTION OFFICIAL REVIEW

DEPARTMENT OF THE ARMY
U.S. ARMY COMBAT CAPABILITIES DEVELOPMENT COMMAND
ARMY RESEARCH LABORATORY
HUMAN RESEARCH & ENGINEERING DIRECTORATE
7101 MULBERRY POINT ROAD
ABERDEEN PROVING GROUND, MARYLAND 21005-5425

CCDC ARL-HRPO 13 Jun 2019

MEMORANDUM FOR: Tamara Griffith, CCDC SC, Orlando, FL

SUBJECT: Human Research Protection Official (HRPO) Review

Protocol/Title: Interactions Between Humans, Virtual Agent Characters and Virtual Avatars (Contract # W911NF-15-2-0003)
Principal Investigator: Tamara Griffith, University of Central Florida, Orlando, FL
Protocol Number: ARL 18-158

Review Outcomes: The CCDC ARL Human Research Protections Official (HRPO) has approved the above referenced project. HRPO review is required for non-DoD institutions conducting research involving human subjects supported by the Department of Defense (DoD), per DoDI 3216.02, Encl 3.4.c.

Requirements:

Substantive Changes to the Protocol: The HRPO must review and accept the IRB's determination when substantive modifications are made to this research study, and any modifications that could potentially increase risk to subjects, before the changes are implemented (DoDI 3216.02, Encl 3.4.c(2)(c)). Substantive changes include: addition of a new institution engaged in the research, change in IRB's determination of review procedure e.g., expedited to fully convened IRB or a change in IRB's Risk Level e.g., minimal to greater than minimal risk.

Significant Changes to the Protocol: The HRPO must review and acknowledge significant changes to the study but these changes may be implemented prior to HRPO review. Significant changes are major non-administrative changes such as a change in project design or change in the institution's Principal Investigator.

Continuing Review: The HRPO must ensure an appropriate continuing review occurred within the required timeframe, (DoDI 3216.02, Encl 3.4.c(2)(d)). Submit communication from the IRB regarding any lapse in IRB approval.

Study Closure: The HRPO must be informed of the date and reason for study closure (i.e., study completed, insufficient enrollment to sustain the research, etc.). The HRPO must receive the final study report submitted to the IRB, including a copy of any acknowledgement documentation and any supporting documents, as soon as all documents become available.

Notification: The investigator must immediately notify the HRPO of the occurrence of any of the following (DoDI 3216.02, Encl 3.4.b(4)):

When the IRB used to review and approve the research changes to a different IRB.

- The knowledge of any pending compliance inspection/visit by the Food and Drug Administration (FDA), Office for Human Research Protections, or other government agency concerning this research; the issuance of inspection reports, FDA Form 483, warning letters, or actions taken by any regulatory agencies including legal or medical actions.

- Suspension or termination of this research study by the IRB, the institution, the sponsor, or regulatory agency.

- Confirmed unanticipated problems involving risks to subjects or others related to this research study.

- Confirmed serious or continuing noncompliance related to this research study.

Caution: Do not construe this correspondence as approval for any contract funding. Only the Contracting Officer/Grants Officer can authorize expenditure of funds. Also, do not construe this as IRB approval.

The POC for technical questions regarding this report is Ms. Jennifer Biladeau at usarmy.apg.ccdc.mbx.arl-irb-office@mail.mil.

STRAUT.THERESA.M.1501857242 Digitally signed by STRAUT.THERESA.M.1501857242 Date: 2019.06.13 15:25:29 -04'00'

THERESA M. STRAUT
Human Research Protection Official, ARL

APPENDIX M - SOCIAL PHOBIA INVENTORY AGREEMENT

Dear Tami:

Thank you for your interest in the Social Phobia Inventory (SPIN). By this agreement you are granted permission to use the scale in the activity you have described, under the following terms:

1. You agree not to provide the scale to a third party unconnected with your project. If other off-site collaborators are involved with your work, their use of the scale is restricted to this work, and the signatory of this agreement is responsible for ensuring that all collaborators adhere to the terms of this agreement.

2. You may use the SPIN in written format for completion as a hard copy, or through administration over the telephone. The SPIN may also be administered in a secure electronic format if arrangements have been made to protect the scale from unauthorized distribution or the possibility of modification. **In all presentations of the SPIN, including electronic versions, the full copyright and terms of use statement must appear with the scale. The SPIN should not appear in any form where it is accessible to the public and should be removed from electronic and other sites once the project has been completed.**

3. The scale's content may not be modified, although in some circumstances the formatting or presentation may be adapted, with permission of Dr. Davidson, after reviewing any proposed adaptations. It is important that the entire copyright statement be retained *verbatim* on all formats of the scale.

4. If you create a non-English language or culturally modified version of the SPIN, please e-mail a copy of the English back translation of the SPIN for review prior to implementing the scale in your work. In addition, please include the following language at the end of the form:

Scale is based upon the English language version of the Social Phobia Inventory, © 1995, 2014, Jonathan R. T. Davidson, MD. All Rights Reserved. Translation by.........

5. For use of the SPIN a student-rate fee of $ 30 US is requested, payable to Jonathan Davidson at 2434 Racquet Club Drive, Seabrook Island, SC 29455, USA, by PayPal (account mail@cd-risc.com at www.paypal.com), cheque, bank wire, international money order or Western Union.

6. Complete and return this form via email to mail@cd-risc.com.

7. In any publication or report resulting from use of the SPIN, you do not publish or partially reproduce the scale.

If you agree to the terms of this agreement, please email a signed copy to the above email address. Upon receipt of the signed agreement and payment, we will email a copy of the scale. For questions regarding use of the SPIN, please contact Jonathan Davidson, at mail@cd-risc.com.

Sincerely yours,

Jonathan R. T. Davidson, M.D.

Agreed to by:

____________________ ___1/30/2020______

Signature Date

Tamara S. Griffith__
Name (printed) (optional)

Chief Engineer, PhD Candidate
Title

U.S. Army, Dissertation through University of Central Florida
Organization

LIST OF REFERENCES

Abhang, P. A., Gawali, B. W., & Mehrotra, S. C. (2016). *Introduction to EEG- and Speech-Based Emotion Recognition.* Academic Press.

Advanced Brain Monitoring, Inc. (2015). *B-Alert X10 EEG Headset System.*

Aitamurto, T., Boin, J.-B., Chen, K., Cherif, A., & Shridhar, S. (2018). The Impact of Augmented Reality on Art Engagement: Liking, Impression of Learning, and Distraction. *International Conference on Virtual, Augmented and Mixed Reality* (pp. 153-171). Springer.

Alexander, A. L., Brunye, T., Sidman, J., & Weil, S. A. (2005). From gaming to training: A review of studies on fidelity immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games. *DARWARS Training Impact Group*, 1-14.

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production; some gestures are meant to be seen. *Journal of Memory and Language*, 44,169-188.

Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: we think, therefore we gesture. *Language and Cognitive Processes*, 15(6), 593-613.

Appel, J., von der Putten, A., Kramer, N. C., & Gratch, J. (2012). Does humanity matter? Analyzing the importance of social cues and Perceived Agency of a computer system for the emergence of social reactions during human computer interaction. *Advances in Human-Computer Interaction.*

Artstein, R., Traum, D., Boberg, D., Gainer, A., Gratch, J., & Leuski, A. (2017). Listen to my Body: Does Making Friends Help Influence People? *Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Confence*, (pp. 430-435). Florida.

Ascoli, G. A. (2014). Neurocognitive models of sense-making. *Biologically Inspired Cognitive Architectures*, (pp. 82-89).

Banks, J., & Bowman, N. D. (2013). Close intimate playthings? Understanding player-avatar relationships as a function of attachment, agency and intimacy. *AoIR Selected Papers of Internet Research*.

Bates, J. (1994). The role of emotion in believable agents. *ACM*, 122-125.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. (2005). The Iowa Gambling Task and the somatic marker hypothesis: some questions and answers. *Trends in cognitive sciences*, 159-162.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and economic behavior*, 122-142.

Berry, P. M., Donneau-Golencer, T., Duong, K., Gervasio, M., Peintner, B., & Yorke-Smith, N. (2017). Evaluating intelligent knowledge systems: experiences with a user-adaptive assistant agent. *Knowledge and Information Systems* (pp. 379-409). Springer.

Blascovich, J. (2002). A theoretical model of social influence for increasing the utility of collaborative virtual environments. *Proceedings of the 4th international conference on Collaborative virtual environments.* ACM.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry, 13*, pp. 103-124.

Blinde, L. (2016, May 17). Army posts solicitation for intelligence & electronic warfare tactical proficiency trainer. *Intelligence Community News*. Retrieved March 20, 2020, from https://intelligencecommunitynews.com/army-posts-solicitation-for-intelligence-electronic-warfare-tactical-proficiency-trainer/

Branigan, H. P., Pickering, M. J., Pearson, J., & McLean, J. F. (2010). Linguistic alignment between people and computers. *Journal of pragmatics 42(9)*, 2355-2368.

Brimstin, J. A., Kemper, B., Parrish, R., Johnston, J., Garrity, P., Moon, J., . . . Butler, P. V. (2015). *Stress exposure training for the dismounted squad: The human dimension.* US Army, Maneuver Center of Excellence, Ft. Benning, United States Army.

Bukhari, H., Andreatta, P., Goldiez, B., & Rabelo, L. (2017). *A framework for determining the return on investment of simulation-based training in health care.* Inquiry. doi:10.1177/0046958016687176

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional system. *Nature Reviews Neuroscience*, 186-202.

Camerer, C. F. (2003). *Behavioral Game Theory.* Princeton, NJ: Princeton University Press.

Chalmers, J. A., Quintana, D. S., Maree, J., Abbott, A., & Kemp, A. H. (2014). Anxiety disorders are associated with reduced heart rate variability: A meta-analysis. *Frontiers in Psychiatry*, 5.

Chang, C.-C., Liang, C., Chou, P.-N., & Lin , G.-Y. (2017). Is game-based learning better in flow experience and various types of cognitive load than non-game based learning? Perspective from multimedia and media richness. *Computers in Human Behavior*, (pp. 218-227).

Connor, K. M., Davidson, J. R., Churchill, I. E., Sherwood, A., Weisler, R. H., & Foa, E. (2000). Psychometric properties of the social phobia inventory (SPIN) new self-rating scale. *The British Journal of Psychiatry*, 176 (40), 379-386.

Cromwell, H. C., Tremblay, L., & Schultz, W. (2013). Neural encoding of choice during a delayed response task in primate striatum and orbitofrontal cortex. *Experimental brain research*, (pp. 1679-1688).

Csikszentmihalyi, M. (1998). *Finding Flow: The Psychology of Engagement with Everyday Life.* New Hork: Basic Books.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. In *Handbook of psychophysiology* (pp. 200-223).

de Melo, C. M., & Gratch, J. (2015). Beyond believability: quantifying the differences between real and virtual humans. *International Conference on Intelligent Virtual Agents* (pp. 109-118). Springer.

de Melo, C., Carnevale, P., & Gratch, J. (2014). Bridging the gap between human and non-human decision makers. *Annual meeting of the international association for conflict management.* Leiden, the Netherlands: .

De Quervain, D. J., Fishbacker, U., Treyer, V., & Schellhammer, M. (2004). The neural basis of altruistic punishment. *Science*, 1254.

Deleuze, G. (1988). *Spinoza: practical philosophy.* City Lights Books.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 9-21.

Disney Movies. (2017). *Official Disney Website.* Retrieved from Disney Movies - Coco: https://movies.disney.com/coco

Drachen, A., Nacke, G., Yannakakis, G., & Pedersen, A. L. (2010). Correlation between Heart Rate, Electrodermal Activity, and Player Experience in First Person Shooter games. *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games,* (pp. 49-54). New York, NY.

Dukstein, G., Watkins, J., & Deakins, R. (2007). Integration of OneSAF Environment Runtime Component into Existing Virtual Simulations. *Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)* (p. Volume 2007). Orlando: NTSA.

Egan, D., Brennan, S., Barrett, J., Qiao, Y., Timmerer, C., & Murray, N. (2016). An evaluation of Heart Rate and Electrodermal Activity as an Objective QoE Evaluation method for Immersive Virtual Reality Environments. *Eighth International Conference on Quality of Multimedia Experience (QoMEX)* (pp. 1-6). IEEE.

Fehr, E., & Fisxhbacker, U. (2003). The nature of human altruism. *Nature*, 785-791.

Feinstein, A. H., & Cannon, H. M. (2014). Fidelity, Verifiability, and Validity of Simulation: Constructs for Evaluation. *Developments in Business Simulation and Experiential Learning*, 58-67.

Fox, J., Ahn, S., Janssen, J. H., Yeykelis, L., Segovia, K. Y., & Bailenson, J. N. (2015). Avatars versus agents: a meta-analysis quantifying the effect of agency on social influence. *Human Computer Interaction*, (pp. 401-432).

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London: Biological Sciences*, 459-473.

Garnham, A. (2017). *Artificial Intelligence: An Introduction.* Routledge.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, *46(2)*, pp. 107-119. doi:http://dx.doi.org/10.1037/0003-066X.46.2.107

Gintis, H. (2014). *The bounds of reason: Game theory and the unification of the behavioral sciences.* Princeton University Press.

Grant, T. (2005). Unifying planning and control using an OODA-based architecture. *Proceedings of the 2005 annual research conference of the South African institute of computer scientists and* *information technologiests on IT research in developing countries.* South African Institute for Computer Scientists and Information Technologists.

Gray, H., Gray, K., & Wegner, D. (2007). Dimensions of mind perception. *Science*, 315,619.

Gregersen, A., Langkjaer, B., Heiselberg, L., & Wieland, J. L. (2017). Following the viewers: Investigating television drama engagement through skin conductance measurements. *Poetics*, 1-13.

Greswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, Fifth Edition.* Thousand Oaks, CA: Sage Publications.

Grey, J., & Bryson, J. J. (2011). Procedural quests: A focus for agent interaction in role-playing-games. *Proceedings of the AISB 2011 Symposium: AI & Games.* Bath, UK: University of Bath.

Guth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of economic behavior & organization*, 367-388.

Halcomb, E. J., & Hickman, L. (2015). Mixed methods research. *Nursing Standard: promoting excellence in nursing care, 29(32)*, 41-47.

Harbaugh, W. T., Mayr, U., & Burghart, D. R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science*, 1622-1625.

Hartanto, D., Kampmann, I. L., Morina, N., Emmelkamp, P. G., Neerincx, M. A., & Brinkman, W. (2014). Controlling Social Stress in Virtual Reality Environments. *PloS One.*

Headquarters, Department of the Army. (2015). *Intelligence Support to Urban Operations.* Washington, DC: U.S. Army.

Headquarters, Department of the Army. (2018). *Army Regulation (AR) 350-50.* Washington, DC: Headquarters, Department of the Army.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, 243-259.

Hertz, J., Krogh, A., & Palmer, R. G. (2018). *Introduction to the Theory of Neural Computation.* Boca Raton, FL: Taylor & Francis Group.

Heyselaar, E., Hagoort, P., & Segaert, K. (2017). In dialogue with an avatar, language behavior is identical to dialogue with a human partner. *Behavior research methods*, (pp. 46-60).

Hossain, Z., Gedeon, T., & Sankaranarayana, R. (2016). Observer's Galvanic Skin Response for Discriminating Real from Fake Smiles. *Australasian Conference on Information Systems.*

HQ Department of the Army. (2014). *Training Circular (TC) 7-102 Operational Environment and Army Learning.* U.S. Army.

HRV Course. (2016). *What's the difference between heart rate and heart rate variability?* Retrieved November 16, 2018, from https://hrvcourse.com/heart-rate-variability-vs-heart-rate/

Ismail, W. W., Hanif, M., Mohamed, S. B., Hamzah, N., & Rizman, Z. I. (2016). Human emotion detection via brain waves study by using electroencephalogram (EEG). *International Journal on Advanced Science, Engineering and Information Technology*, 1005-1011.

Jacobs, N., & Garnham, A. (2006). The role of conversational hand gestures in a narrative task. *Jornal of Memory and Language*, 291-303.

Kazmi, S. Z., Zhang, H., Aziz, W., Monfredi, O., Abbas, S. A., Shah, S. A., . . . Butt, W. H. (2016). Inverse Correlation between Heart Rate Variability and Heart Rate Demonstrated by Linear and Nonlinear Analysis. *PloS One*, 11(6).

Krauss, R. M., Dushay, R. A., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology*, 31,533-552.

Lacity, M., Willcocks, L., & Craig, A. (2017). Service Automation: Cognitive Virtual Agents at SEB Bank. *The Outsourcing Unit Working Research Paper Series* (pp. 1-29). London School of Economics and Political Science.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 1-3.

Langer, E. J. (1992). Matters of mind: mindfulness/mindlessness in perspective. *Consciousness and Cognition*, (pp. 289-305).

Langer, E. J., & Moldoveanu, M. (2000). The construct of mindfulness. *Journal of Social Issues*, (pp. 1-9).

Lee, E. J., Nass, C., & Brave, S. (2000). Can computer-generated speech have gender? An experimental test of gender stereotypes. *CHI 2000 Conference Companion.* New York: Association of Computing Machinery.

Li, M., Jiang, Q., Tan, C.-H., & Wei, K.-K. (2014). Enhancing user-game engagement through software gaming elements. *Journal of Management Information Systems, 30*(4), 115-150.

Lim, S., & Reeves, B. (2010). Computer agents versus avatars: Responses to interactive game characters controlled by a computer or other player. *International Journal of Human-Computer Studies*, *68*, pp. 57-68.

Maren, A. J., Harston, C. T., & Pap, R. M. (1990). *Handbook of Neural Computing Applications.* San Diego, CA: Academic Press, Inc.

Maxwell, D. B., Griffith, T. S., & Finkelstein, N. (2014). The Use of Virtual Worlds in the Military Services as Part of a Blended Learning Strategy. In K. S. Hale, & K. M. Stanney, *Handbook of*

Virtual Environments: Design, Implementation, and Applications, Second Edition (pp. 959-999). Orlando, FL: CRC Press.

McCabe, K., Houser, D., Ryan, L., Smith, Vernon, & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences*, (pp. 11832-11835).

Merriam-Webster. (2019). Kinesics. *Merriam-Webster.com*. Retrieved September 23, 2019, from www.merriam-webster.com/dictionary/kinesics

MidBrain Power. (2014). *Parts and Functions of Human Brain.* Retrieved from Technical Details: www.midbrainpower.in/technical-details/supersonic-human-brain/parts-and-functions-of-human-brain

Milham, L. M., Phillips, H. L., Ross, W. A., Townsend, L. N., Riddle, D. L., Smith, K. M., . . . Johnston, J. H. (2017). Squad-level training for Tactical Combat Casualty Care: Instructional Approach and Technology Assessment. *The Journal of Defense Modeling and Simulation*, 345-360.

Millington, I., & Funge, J. (2009). *Artificial Intelligence for Games.* CRC Press.

Moon, Y. (1998). *Intimate self-disclosure exchanges: Using computers to build reciprocal relationships with consumers.* Boston: Harvard Business School.

Moon, Y., & Nass, C. (1996). How "real" are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication Research*, *23*, pp. 651-674.

Mor, S., Cortez, P., & Rita, P. (2018). A divide-and-conquer strategy using feature relevance and expert knowledge for enhancing a data mining aproach to bank telemarketing. *Expert Systems.*

Mori, M. (1970). The uncanny valley. *Energy*, 33-35.

Morkes, J., Kernal, H., & Nass, C. (1999). Effects of humor in task-oriented human-computer interaction and computer-mediated communication: A direct est of SRCT theory. *Human-Computer Interaction*, *14(4)*, pp. 395-435.

Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similiarity-attraction, and consistency-attraction. *Journal of experimental psychology*, 171.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues, 56.1*, 81-103.

Nass, C., Moon, Y., & Carney, P. (1999). Are respondents polite to computers? *Journal of Applied Social Psychology, 29(5)*, pp. 1093-1110.

Nass, C., Tauber, E., & Reeder, H. (1993). Anthropomorphism, agency and ethopoeia: computers as social actors. *Proceeding of the International Conference Companion on Human Factors in Computing Systems* (pp. 111-112). CHI '93.

Neulog . (2018). *NeuLog Galvanic Skin Response Sensor User Guide.*

Niedermeyer, E., & da Dilva, F. L. (2005). *Electroencephalography: basic principles, clinical applications, and related fields.* Lippincott Williams & Wilkins.

Noordzij, M. L., Scholten, P., & Laroy-Noordzij, M. E. (2012). Measuring electrodermal activity of both individuals with severe mental disabilities and their caretakers during episodes of challenging behavior. In A. Spink, F. Grieco, O. Krips, L. Loijens, L. Noldus, & P. Zummerman (Ed.), *Proceedings of Measuring Behavior 2012* (pp. 201-205). Utrecht, The Netherlands: Wageningen, The Netherlands: Noldus.

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 481-494.

Oosterbeek, H., Sloof, R., & van de Kuilen, G. (2003). Cultural differences in ultimatum game experiements: Evidence from a meta-analysis. *Experimental Economics* (pp. 171-188). Springer Science and Business Media.

Ozyurek, A. (2002). Do speakers design their cospeech gesture for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language*, 46, 688-704.

Paiva, A. (2000). *Affective interactions: Toward a new generation of computer interfaces?* Berlin: Springer.

Pearce, J. (2005). Engaging the Learner: How can the flow experience support e-learning? *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, (pp. 2288-2295).

Pena, J., Khan, S., & Alexopoulos, C. (2016). I am what I see: How avatar and opponent agent body size affects physical activity among men playing exergames. *Journal of Computer-Mediated Communication*, (pp. 195-209).

Poundstone, W. (1993). Prisoner's Dilemma. In J. von Neumann, *Game Theory and the Puzzle of the Bomb.* Anchor.

Pradana, A., Goh, O. S., & Kumar, Y. J. (2018). Intelligent Conversational Bot for Interactive Marketing. *Journal of Telecommunication, Electronic and Computer Engineering*, 1-7.

Rafaeli, S. (1985). *Interacting with media: Para-social interaction and real interaction, Doctoral Dissertation.* Stanford University.

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge University Press.

Ricard, G., & Warzynski, F. (2014). Vertical integration, exclusivity, and game sales performance in the US video game industry. *Journal of Law, Economics and Organization.*

Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annual review of psychology*, (pp. 23-48).

Rilling, J. K., Gutman, D. A., Pagnoni, G., & Berns, G. S. (2002). A neural basis for social cooperation. *Neuron*, (pp. 395-405).

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *Neuroimage*, 1694-1703.

Rime, B. (1982). The elimination of visible behavior from social interactions: effects on verbal, non-verbal and interpersonal variables. *European Journal of Social Psychology*, 12,113-129.

Ritterfeld, U., Cody, M., & Vorderer, P. (2009). *Serious games: Mechanisms and effects.* Routledge.

Rossini, P. M., Burke, D., Chen, R., Cohen, L. G., Daskalakis, Z., Di lorio, R., . . . Ziemann, U. (2015). Non-invasive electrical and magnetic stimulation of the brain, spinal cord, roots and peripheral nerves: Basic principles and procedures for routine clinical and research application. An updated report from an I.F.C.N Committee. *Clinical Neurophysiology*, 126.6 1071-1107.

Russell, S. J., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach.* Malaysia: Pearson Education Limited.

Sanfey, A. G. (2007). Social decision-making: Insights from Game Theory and Neuroscience. *Science*, (pp. 598-602).

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science* , 1755-1758.

Schobel, S., Janson, A., Hopp, J. C., & Leimeister, J. M. (2019). Gamification of Online Training and its Relation to Engagement and Problem-solving Outcomes. Boston, Massachusetts: Academy of Management Annual Meeting.

Shernoff, D. J., Csikszentmihalyi, M., Schneider, B., & Shernoff, E. S. (2014). Student engagement in high school classrooms from the perspective of flow theory. *Application of flow in human development and education* (pp. 475-494). Dordrecht: Springer.

Singer, T., Seymour, B., O'doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, 439-466.

Smith, R. (2010). The Long History of Gaming in Military Training. *Simulation Gaming*, vol 41, no 1, 6-19.

Stephenson, N. (1995). *The Diamond Age: Or, a Young Lady's Illustrated Primer.* Bantam Books.

Summers, J. E. (2012). Simulation-based Military Training: An Engineering Approach to Better Addressing Competing Environmental, Fiscal and Security Concerns. *J. Wash. Acad. Sci.*, 9-22.

Sundar, S. S. (1993). Communicator, programmer, or independent actor: What are computers? *Annual conference of the International Communication Association.* Washington, DC.

Sundar, S. S., & Nass, C. (2000). Source Orientation in Human-Computer Interaction: Programmer, Networker, or Independent Social Actor? (I. Sage Publication, Ed.) *Communication Research, 27(6)*, 683-703.

Teplan, M. (2002). Fundamentals of EEG Measurement. *Measurement Science Review*, Volume 2, Section 2, 1-11.

Thomas, F., & Johnston, O. (1981). *Disney Animation: The Illusion of Life.* New York: Abbeville Press.

Trans Cranial Technologies. (2012). *10/20 System Positioning Manual*. Hong King: Trans Cranial Technologies, Ldt.

Turkle, S. (1984). *The second self: Computers and the human spirit.* New York: Simon and Schuster.

U.S. Army Combined Arms Center. (2014). *The Human Dimension White Paper: A Framework for Optimizing Human Performance*. U.S. Army.

U.S. Army Modeling and Simulation Office. (2011, June 30). *Army Modeling and Simulation Office of Deputy Chief of Staff, G-8, CAA*. Retrieved August 19, 2011, from History of Modeling and Simulation (M&S): http://www.ms.army.mil/about/history/index.html

U.S. Army Modeling and Simulation Office. (2016, January 6). *STAND-TO*. Retrieved from AMSO: https://www.army.mil/standto/archive_2016-01-06

U.S. Army Training and Doctrine Command. (2014). *The Army Operating Concept.* Fort Eustis, VA: U.S. Government Printing Office.

Vice, J., Skinner, A., Berka, C., Reinerman-Jones, L., Barber, D., Pojman, N., . . . Lathan, C. (2011). Use of Neurophysiological Metrics within a Real and Virtual Perceptual Skills Task to Determine Optimal Simulation Fidelity Requirements. *Virtual and Mixed Reality - New Trends - Lecture Notes in Computer Science* (pp. Volume 6773; pp 387-399). Springer.

Vogel, R. (2016). *Overview of EEG.* Retrieved August 30, 2018, from https://neurologiclabs/neuromonitoring/eeg

von der Putten, A. M., Kramer, N. C., Gratch, J., & Kang, S. H. (2010). It doesn't matter what you are! Explaining social effects of agents and avatars. *Computers in Human Behavior*, (pp. 1641-1650).

Webster, A. (2016, Sept 14). *In these games starring AI bots, chatting is the new shooting.* Retrieved from The Verge: https://www.theverge.com/2016/9/14/12905356/komrad-evento-chatbot-games-preview-conversation-ai-personality

White, N. E., & Richards, L. M. (2009). Alpha-theta neurotherapy and the neurobehavioral treatment of addictions, mood disorders and trauma. In T. H. Budzynski, H. K. Budzynski, J. R. Evans, & A. Abarbanel, *Introduction to Quantitative EEG and Neurofeedback: Advanced Theory and Applications* (pp. 143-164). Elsevier.

Wilson, G., & Sasse, M. A. (2000). Listen to your heart rate: Counting the cost of media quality. In A. Paiva, *Affective interactions: Toward a new generation of computer interfaces* (pp. 9-20). Berlin: Springer.

Winograd, T., & Flores, C. (1987). *Understanding computers and cognition: A new foundation for design.* Reading, MA: Addison-Wesley.

Wisdom, J. P., Cavaleri, M. A., Onwuegbuzie, A. J., & Green, C. A. (2012). Methodological reporting in qualitative, quantitative and mixed methods health services research articles. *Health Services Research*, 721-745.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, (pp. 225-240).

Yokotani, K., Takagi, G., & Wakashima, K. (2017). Advantages of virtual agents over clinical psychologists during comprehensive mental health interviews using a mixed methods design. *Computers in Human Behavior*.

Zang, W., & Creswell, J. (2013). The use of "mixing" procedure of mixed methods in health services research. *Medical Care*, e51-7.